THE UNIVERSITY
of EDINBURGH

BIOCHEMICAL AND BIOPHYSICAL INVESTIGATIONS INTO KEY MALARIA PARASITE PROTEINS.

Christopher Neil Haggarty-Weir

ORCID: https://orcid.org/0000-0001-9846-6287

Submitted in total fulfilment of the requirement of the degree of Doctor of Philosophy (Ph.D)

EaStChem School of Chemistry, College of Science and Engineering at The University of Edinburgh,

and the

Department of Medical Biology, Faculty of Medicine at The University of Melbourne

September 2017

Scotland and Australia.

Abstract

Plasmodium falciparum, the most pestilential of the malaria parasite species, is responsible for ~450,000 direct deaths annually. Clinical disease is a consequence of the blood stage of the parasite's lifecycle involving a plethora of host-parasite interactions. Key to these interactions are the *P. falciparum* reticulocyte binding-like homologue (PfRh) proteins responsible for binding erythrocyte receptors and gaining entry to host cells. For example, PfRh4 binds to human complement receptor-1 (CR1) on erythrocytes for sialic-acid-independent invasion. Another protein important for invasion is the PfRh5-interacting protein (PfRipr), an essential member of the PfRh5-associated invasion complex (PAIN-complex) along with CyRPA, the cysteine-rich protective antigen. Loss of function of PfRipr in *P. falciparum* parasites prevents erythrocyte entry and ablates Ca^{2+}-influx into the erythrocyte; essential events during invasion.

This study aimed to biochemically and structurally investigate truncated recombinant versions of PfRh4 and PfRipr. Homology modelling suggested that PfRh4 is rich in alpha-helical secondary structure. The sequence of PfRipr suggested the presence of ten epidermal growth factor-like (EGF) modules, two towards the N-terminus and eight in the C-terminal domain.

In this project, monoclonal antibodies made against recombinant PfRh4 were shown, *via* indirect immunofluorescent assays, to localize to the apical tip of merozoites. Monoclonal antibody 5H12, raised against PfRh4, reduces parasite invasion of erythrocytes by ~75% in growth-inhibition assays with neuraminidase pre-treated erythrocytes. Attempts to produce a stable truncated recombinant PfRh4 protein for structural studies were unsuccessful. An ELISA-based assay using ten alanine-scan mutants suggested the CR1-binding site lies outside of amino acids 283 – 341 of PfRh4.

PfRipr truncations, defined by the boundaries of EGF-like repeats predicted based on sequence homology, were produced recombinantly in *Escherichia coli* and *Pichia pastoris*. These proteins had a circular dichroism signature suggestive of β-strand-containing proteins with disordered regions. EGF-containing PfRipr truncations did not bind recombinant PfRh5 according to ELISA and size-exclusion chromatography assays. EGFs 1-2, 5-7 and 7-10 of PfRipr did not bind CyRPA via size-exclusion chromatography or NMR. Crystallisation trials performed on EGF modules failed to yield crystals suitable for data collection. A ^{15}N isotopically-labelled sample of EGF5-7 gave good quality HSQC NMR spectra. A suite of three-dimensional NMR spectra collected on a ^{13}C,^{15}N-EGF5-7 sample, at three different temperatures, allowed for >86% of backbone assignments. T_1/T_2 relaxation analysis and heteronuclear NOE data were suggestive of an elongated, rigid protein undergoing intermolecular self-association. Further evidence for EGF5-7 being an elongated protein was provided via SAXS analysis. Chemical shifts facilitated prediction of secondary structure in EGF 5-7 consistent with an EGF-like fold. Melting studies performed on EGF5-7 showed no evidence of denaturation over the temperature range 20 °C - 95 °C indicating a thermally-stable protein. The addition of Ca^{2+} to the ^{15}N-EGF5-7 sample caused chemical shift perturbations consistent with high-affinity binding. The discovery of inhibitory monoclonal antibodies recognising a conformational epitope on EGF7 provided evidence of the functional importance of this region within PfRipr.

The work described in this thesis provides methods for the industrially-scalable production and biophysical investigations of *P. pastoris* or *E. coli*-produced disulfide-rich *P. falciparum* antigens of interest to vaccinologists.

Acknowledgements

First and foremost, I must thank my dear wife Stephanie Haggarty-Weir, in addition to my wonderful mother Karen Weir; both of whom have given me enormous support and strength to be able to overcome the numerous personal, professional and medical obstacles I have found myself confronted with during my doctoral research.

Secondly, I want to thank the students I have trained who contributed positively to my professional and personal life, Roma Galloway and Ella Svahn. Their intellectual contributions have been highly valued and seeing their professional development has been immensely rewarding.

Thirdly, I want to express my enormous gratitude to the staff and colleagues who have gone above and beyond to provide me with valuable assistance in my studies. In Australia, this includes Joan Curtis, Dr. Melissa Call, Sue Hardy and William Godfrey (for helping me with last minute data analysis and figure generation). In Scotland, this includes Juraj Bella, Dr. John White, Dr. Sarah Reece, Dr. Dave Clark and Denise Wilson. Appreciation should also be given to my thesis advisory committee, with particular mention given to Prof. Alan Cowman who helped me secure additional funding to complete my Ph.D, Dr. Anthony Hodder for his hands-on training of me, Dr. Keely Bumsted-O'Brien for being someone I was always comfortable talking to, and to both A/Prof. Matthew Call and Prof. Dusan Uhrin for their personal and professional guidance at important points of my Ph.D.

I wish to also give thanks to Dr. Michael Diamond, and the Disability Service team plus the student counselling team of the University of Edinburgh who provided me with enormous help when I was undergoing significant mental health crises. Without their help I wouldn't be here today.

Lastly, tribute must be paid to all my funders, without whom none of this would have been possible. This includes the Australian Society for Parasitology and the Australia - Europe Malaria Research Cooperation (OzEMalaR) for my travel grant, all 86 backers of the Pozible crowdfunding project who helped fund the protein refolding work and some of the NMR research, and the Australian and British taxpayers who funded my research and grants.

Declaration

Unless otherwise specifically stated in the text, all the work given in this thesis is that of my own and has not been submitted whole or in part for a degree or any other form of qualification. This includes methodologies. Work described in this thesis has been published in Nicholas T.Y. Lim, Markus J. Harder, Alexander T. Kennedy, Clara S. Lin, Christopher Weir, Alan F. Cowman, Melissa J. Call, Christoph Q. Schmidt and Wai-Hong Tham. Characterization of inhibitors and monoclonal antibodies that modulate the interaction between *Plasmodium falciparum* adhesin PfRh4 with its erythrocyte receptor Complement Receptor 1. *J. Biol. Chem.* 2015, 290:25307-25321. This paper is included at the end of the appendix to this thesis.

Additional work described in this thesis has been submitted for publication and is under review in Christopher N. Haggarty-Weir, Roma Galloway, William Godfrey. Methods for the Refolding of Disulfide-Rich Proteins. *EUREKA! Methods* 2017 (currently under preprint on the BioArxiv server- http://www.biorxiv.org/content/early/2017/06/30/158162, and as a peer review-pending copy on Zageno's *EUREKA! Methods* page- https://zageno.com/e/refolding-disulfide-rich-proteins). The submitted manuscript is also included at the end of the appendix to this thesis.

Contents

Abbreviations

1D	One-dimensional.
2D	Two-dimensional.
3D	Three-dimensional.
Abs_{280}	Absorbance at 280 nm.
ACN	Acetonitrile.
AMA-1	Apical membrane antigen 1.
AOX	Alcohol oxidase.
BCE	B-cell epitope.
BMG	Buffered minimal glycerol.
BMGY	Buffered complex medium containing glycerol.
BMM	Buffered minimal methanol.
BMMY	Buffered complex medium containing methanol.
BSA	Bovine serum albumin,
BSG	Basigin.
Ca^{2+}	Cationic calcium.
CCP	Complement control protein.
coIP	Co-immunoprecipitation.
CR1	Complement receptor 1.
CSP	Circumsporozoite protein.
CyRPA	Cysteine-rich protective antigen.
DAB	3,3`-diaminobenzidine.
DIV	Direct Intravenous Route.
DNA	Deoxyribonucleic acid.
DTT	Dithiothreitol.
EBA	Erythrocyte binding antigen.
EBL	Erythrocyte binding ligand.
EDTA	Ethylenediaminetetraacetic acid.
EDV	Electron-dense granules.
E. coli	*Escherichia coli.*

EGF	Epidermal growth factor.
EK	Enterokinase.
ELISA	Enzyme-linked immunosorbent assay.
EPPF	Edinburgh Protein Production Facility.
ESI-MS	Electrospray ionization mass spectrometry.
FRET	Förster resonance energy transfer.
GAP	Genetically-attenuated parasite.
GIA	Growth inhibition assay.
GDP	Gross Domestic Product.
GMQE	Global Model Quality Estimation.
GPI	Glycophosphatidylinositol.
HPLC-ESI/MS	High performance liquid chromatography – Elecrospray ionization/mass spectrometry.
HRP	Horse radish peroxidase.
HSPG	Heparin sulphate proteoglycans.
HSQC	Heteronuclear single quantum coherence.
I	Induced
IFA	Immunofluorescent assay.
IgG	Immunoglobulin G.
IPTG	Isopropyl β-D-1-thiogalactopyranoside.
KAHRP	Knob-associated histidine-rich protein.
kDa	Kilodaltons.
LHR	Long homologous repeat.
LC-MS	Liquid chromatography-mass spectrometry.
mAb	Monoclonal antibody.
MCT	Monocarboxylic acid transporter.
MES	2-(N-morpholino)ethanesulfonic acid.
ms	Milliseconds.
MMV	Medicines for Malaria Venture.
MOPS	(3-(N-morpholino)propanesulfonic acid).

MSP	Merozoite surface protein.
MTRAP	Merozoite thrombospondin-related anomalous protein.
MW	Molecular weight
NA	Neuraminidase.
NEM	N-ethylmaleimide.
NI	Non-induced.
Ni-NTA	Nickel nitrilotriacetic acid.
NMR	Nuclear Magnetic Resonance.
NOE	Nuclear Overhauser effect.
NOESY	Nuclear Overhauser effect spectroscopy.
NPP/PSC	New permeability pathways.
NR	Non-reducing.
NRMSD	Normalised root mean square deviation.
NRSB	Non-reducing sample buffer.
OD_{600}	Optical density at 600 nm.
PAIN-complex	PfRh5-Associated INvasion-complex.
PBS	Phosphate buffered solution.
P. falciparum	*Plasmodium falciparum.*
PfEMP1	*P. falciparum* erythrocyte membrane protein 1.
PfRh	*P. falciparum* reticulocyte binding-like homologue.
PfSPZ	*P. falciparum* Sporozoite vaccine.
P. pastoris	*Pichia pastoris*.
ProEx	pProExHTA plasmid.
PTEX	*P. falciparum* translocon of exported proteins.
PVM	Parasitophorous vacuole membrane.
R	Reducing.
RESA	Ring-exported surface antigen.
RDC	Residual dipolar coupling.
RNA	Ribonucleic acid.
RON	Rhoptry neck protein.
RP-HPLC	Reverse-phase high-performance liquid chromatography.

RSB	Reducing sample buffer.
SA	Sialic acid.
SAXS	Small angle X-ray scattering.
SB	Sample buffer.
SE	Standard error.
SEC	Size exclusion chromatography.
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis.
SPR	Surface plasmon resonance.
SS. *Pichia*	Super Strain *Pichia*
TCA	Trichloroacetic acid.
TCE	T-cell epitope.
TCEP	Tris(2-carboxyethyl)phosphine.
TEV	Tobacco etch virus enzyme.
TMB	3,3,5,5-tetramethylbenzidine.
TOCSY	Total correlation spectroscopy.
TSA	Thermal shift assay.
USD	United States Dollars.
WHO	World Health Organisation.
YPD	Yeast extract peptone dextrose.
YPDS	Yeast extract peptone dextrose sorbitol.

CHAPTER 1

Introduction: Malaria in the modern world, *Plasmodium falciparum* biology and parasitic biochemistry pertaining to vaccine development.

"*Where malaria prospers most, human societies have prospered least*"

Jeffrey Sachs and Pia Malaney, *Nature* 2002.

"*Satisfaction consists in freedom from pain, which is the positive element of life*"

Arthur Schopenhauer, *Studies in Pessimism* 1891.

1.1 Overview

Malaria is an ancient disease poetically known as "the scourge of humanity", and for good reason- since the first confirmed human case of the disease dating back to 450 A.D. (Sallares & Gomzi, 2001), it is estimated that half of all people who have ever lived have succumbed to this pathogen (Whitfield, 2002). With this ancient disease comes a rich and fascinating history, from mankind originally thinking the disease originated from so-called miasmas (malaria as a word comes from the medieval Italian "mal'aria", meaning "bad air") (Hempelmann & Krafts, 2013), to famous world-changers that are thought to have fallen prey to this parasitic infection such as Alaric I, King of the Visigoths who sacked Rome from 401 A.D (Galassi, *et al*., 2016). Obviously, it is beyond the scope of this thesis to examine the history of malaria (interested parties are directed to (Packard, 2008)), however a brief look into two of the major discoveries in the history of malaria should be warranted so as to set the tone post-overview.

Pertaining to the idea that malaria was caused by some form of toxic air, it was not until the 19th century that the paradigm around the aetiology was shifted thanks to the heroic efforts of two military doctors Charles Alphonse Laveran and Ronald Ross. During the early days of medical microscopy, Laveran made the discovery that blood of malarious patients contained parasites (Laveran, 1893). Initially, he observed small, spherical bodies complete with flagella amongst the blood cells obtained from an infected patient's blood sample. Laveran notes "*I was still hesitating whether these elements were parasites, when on November 6th, 1880, on examining the pigmented spherical bodies mentioned above, I observed, on the edge of several of these elements, moveable filaments or flagella, whose extremely rapid and varied movements left no doubt as to their nature*" (Laveran, 1893). This was of course what we know now to be the gametocyte form of the *Plasmodium* parasite, but Laveran would go on to document various morphological forms. Ronald Ross was the first to demonstrate definitively that mosquitoes were infecting people with malaria, a feat accomplished by two years of microscopically examining thousands of mosquitoes fed with infected

patient blood (Hempelmann & Krafts, 2013) (Ross, 1923). Upon examination of the mosquitoes fed with infected blood (in contrast to the control group), Ross could observe the characteristic pigment that we now know today to be haemozoin, a metabolic by-product generated from *Plasmodium* parasites conversion of the haem encountered in erythrocytes. Due to the lack of pigment observed in the control mosquitoes, it was in 1897 that Ross correctly deduced that mosquitoes must be the agent responsible for the transmission of malaria parasites.

Both Laveran and Ross went on to be awarded the Nobel Prize in Medicine or Physiology (1907 and 1902, respectively) for their paradigm-shifting work. Furthermore, these discoveries are of great importance from both a historical and philosophical stand-point. Inductively, it makes sense that people had thought "bad air" around swamps and marshes led to malaria (or "paludisme" as Laveran preferred, which is from the Latin root for swamp); swamps, of course are breeding grounds for mosquitoes and those in swamp-dense areas can be at a higher risk of contracting malaria (Beck, *et al.*, 1994). However, this is the central philosophical problem with inductive reasoning. Just because swampy areas gave rise to more suffering from malaria, it does not mean the swamp air was the cause; this is a form of *post hoc ergo propter hoc* fallacy in logic and reasoning. What Laveran and Ross were able to do was to take a more scientifically rigorous deductive approach that possessed falsifiable hypotheses and methodologies open to independent verification. This was indeed the case when the world was initially skeptical over the claims made by Laveran; however, once other researchers were also able to observe and report on finding parasitic microorganisms in patient blood the skepticism subsided (Auterhoff, 1967). Interested readers in the philosophy of science, particularly in the superiority of deductive means of reasoning over inductive ones, and paradigm shifts are referred to Popper, 1957, Kuhn, 1962, and Godfrey & Haggarty-Weir, 2017.

The rich history of malaria also encompasses a vast variety of disciplines, from epidemiology to economics, from basic biology to biochemistry, drug/vaccine development and structural biophysical chemistry. It is these fields that shall be examined in more depth with respect to malaria in the introduction to this thesis to provide the reader with an integrated understanding of the field of malariology, before embarking on an in-depth biophysical and biochemical analysis of key *P. falciparum* invasion ligands that are of interest to anti-malaria vaccine development, a development researchers hope will lead to some freedom from parasite-related pain, a truly positive element in any life.

1.2 Malaria: A Global Problem

For as long as there have been *Homo sapiens,* there has been malaria. This stems from the rich evolutionary history of the *Plasmodium* species, which emerged shortly after the appearance of mammals around 200 million years ago (Rich & Ayala, 2006). Whilst there are six species of *Plasmodium* parasites that can cause human disease (*P. falciparum, ovale wallikerii, ovale curtisii* , *vivax, malariae* and *knowlesi*), the most malignant, *P. falciparum*, is thought to have appeared relatively recently in evolutionary history, around 5000 - 10 000 years ago (Rich, *et al.*, 2009), which is consistent with a relatively recent worldwide expansion of *P. falciparum* populations (Joy, *et al.*, 2003). Current genetic and phylogenetic analysis suggests that human *falciparum* malaria is of gorilla origin, with the parasite species making a single cross-transmission from gorilla to human (Liu, *et al.*, 2010). Furthermore, making an equally important contribution to the history of human malaria around the world is the widespread distribution of the disease vector, female *Anopheline* mosquitoes. There are approximately 70 species of *Anopheles* mosquitoes with the capacity to transmit *Plasmodium* parasites (Manguin, *et al.*, 2008) allowing for the presence of endemic malaria in around 109 countries, with a majority found in African, Asian and Latin American regions (Manguin, *et al.*, 2010). Due to the vast

geographic spread of these regions, malaria is truly a global problem. We shall now examine the distribution, spread and economic impact of malaria.

1.2.1 The Distribution and Spread of Malaria

Since 2010, there has been an impressive reduction in global malaria mortality from ~1.24 million (Murray *et al.*, 2012), to approximately 429 000 deaths (Alonso, 2017). This has been attributed to increased access to anti-malarial interventions including more diagnostic testing (such as a 77% increase in paediatric malarial diagnostic testing in sub-Saharan Africa, the most malarious region on earth), a five-fold increase in preventative treatment of pregnant women since 2011, and a doubling in the use of insecticide-treated bed nets (WHO, 2016). However, despite some important successes, professionals and governmental authorities should not rest on their laurels; there were an estimated 212 million new malaria cases in 2015 and currently, every two-minutes another child dies from malaria (Alonso, 2017). Additionally, the World Health Organisation estimates that under half of the 91 malaria-affected countries will meet the upcoming milestone of a 40% reduction in incidence and mortality from the disease by 2020 (WHO, 2016). The current global distribution of malaria is shown in figure 1.1.

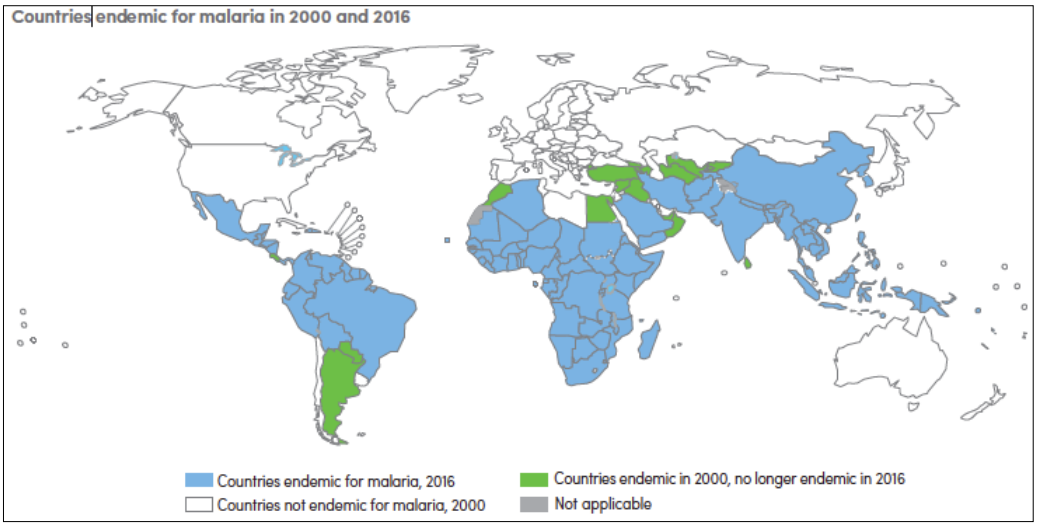


Figure 1.1. The Current Global Distribution of Malaria: Whilst progress has occurred in limiting the spread of endemic malaria (as shown by the countries in green) from 108 countries in 2000 to 91 in 2016, half of the world's population remains at risk (due to the population densities in the blue-coloured countries). This figure also highlights the equatorial focus of malaria endemicity. Figure from (WHO, 2016).

The actual present situation of global malaria morbidity and mortality may indeed be far worse than initially thought. This is due to a combination of under-reporting from malaria surveillance programs (in 2015 for example, it is estimated that these programs only accounted for 19% of global cases) (WHO, 2016), problems with an over-reliance on passive case detection rather than the more rigorous (and expensive) active case detection (Griffin, *et al.*, 2014), and the role of infections secondary to and exacerbated due to malaria. One example of this is the case of the 2008 Mozambique study, where initially 10% of maternal deaths were found to be due to malaria, but 13% were secondary to HIV-infection which can be exacerbated by coexisting malarial infection (Menéndez, *et al.*, 2008).

The spread of malaria is also affected by factors such as climate change (Tanser, Sharp, & Sueur, 2003) and counterfeit drugs (Newton, *et al.*, 2006). Climate change is one of the greatest challenges faced by modern man and has a significant impact on global disease trends (McMichael, Woodruff, & Hales, 2006). With respect to

malaria, modelling indicates a growth in populations at risk due to increased climate suitability for the *Anopheline* mosquitoes that transmit the parasite (Caminade, *et al.*, 2014). Regarding the African continent, one recent decade-long study concluded that climate change was unlikely to have a significant impact in one West African region (with potential reductions in malaria burden) (Yamana, *et al.*, 2016), this may be offset by another 2016 publication concluding a likely increase in tropical, highland and Eastern regions of Africa (Ngarakana-Gwasira, *et al.*, 2016). Whilst there are a lot of uncertainties about the modelling due to alterations in population dynamics, government intervention programs, type of modelling used, alterations of socio-economic factors and changes in the rates of climate change, other studies are in general agreeance with the risk that climate change presents to the spread of malaria and suggest that there is cause for concern at both the local-level and the wider global-level (Ryan, *et al.*, 2015) (Onyango & Oz Sahin, 2016) (Ngarakana-Gwasira, *et al.*, 2016). This should be particularly concerning for the African continent due to it not only having the most endemic levels of malaria transmission but also having a faster than average global warming rate of temperatures, with predictions showing an increase by between 3 and 6 °C by 2100 (Niang, *et al.*, 2014).

Another factor that influences the spread of malaria is that of counterfeit drugs. These can come in one of two major forms; those that consist of sub-clinical doses of active ingredient, and those which have no active ingredient in the compound sold. Both are problematic for slightly different reasons; the former has the potential to lead to the spread of drug-resistant parasites due to the subclinical dose of the drug which may fail to clear parasitaemia, and the latter having the potential to increase mortality/morbidity directly (Karunamoorthi, 2014). It has been estimated that up to half of the anti-malarials in Africa and Southeast Asia are of diminished quality, often due to counterfeits (Karunamoorthi, 2014). One of the most shocking figures was the revelation in 2003 that 88% of all artesunate in the private sector in Laos was fake (Sengaloundeth, *et al.*, 2009). Horrifically,

following Latin hypercube sampling (using estimates calculated from the number of private sector anti-malarials taken by infected children), around 122 350 malaria-related deaths of children under the age of five in Africa during 2013 were deemed likely to be due to low-quality pharmaceuticals (Renschler, *et al.*, 2015).

Regarding fake anti-malarials and their contribution to drug-resistance amongst parasite populations, modelling data does indeed suggest that patients receiving sub-standard treatment are likely to help spread resistance or spontaneous outbreaks of infection (White, *et al.*, 2009). One case study in particular highlights this, the acute 2003 epidemic of malaria on the Pakistan/Afghanistan border was due to substandard locally-produced sulfadoxine/pyrimethamine, with lower bioavailability than the original manufacturers product (Leslie, *et al.*, 2009). It is for the aforementioned reasons that knowingly using, producing or spreading fake anti-malarials should be considered a crime against humanity, and something governments and regulators must take notice of. Those further interested in the use of anti-malarials, the patterns of resistance and the impact on policy, are referred to (Kocovski, Godfrey & Elkington *et al.*, 2015).

1.2.2 The Economics of Malaria

The cost to human life and well-being is rather self-evident; however, this scourge also creates significant socio-economic burdens due to the bi-directional link between malaria and poverty. This is particularly the case in endemic African nations where up to 40% of an individual health budget may be allocated to tackling malaria (Narasimham & Attaran, 2003) (Jowett & Miller, 2005). For example, in 2004 it was calculated that the Tanzanian government spent 1% of its GDP and almost a third of its resources on malaria (Jowett & Miller, 2005). It is important to note that these figures fail to consider the effects on longer-term economic growth and development, in addition to the accounting method used to arise at these figures, which may be unsuitable when estimating economic strain

from areas with a very high malaria transmission rates (Sachs & Malaney, 2002). Furthermore, direct medical intervention including diagnosis and treatment of people infected with malaria only takes into account part of the economic burden; additional costs can also be directed towards preventative measures including purchase of anti-mosquito coils, spraying programs, insecticide-treated bed net distribution, malaria education, epidemiology research, drug quality control and vaccine research and development (Narasimham & Attaran, 2003) (Jowett & Miller, 2005) (Sachs & Malaney, 2002) (Phillips & Phillips-Howard, 1996). Chronic, longer-term strain on budgets additionally includes effects on both microeconomic factors such as the individual household (i.e. migration and saving), and macroeconomic stressors consisting of adverse impacts on trade, foreign investment and tourism (Sachs & Malaney, 2002). One worrisome figure, if many of the aforementioned factors (including life expectancy, location, economic policy and initial poverty) are taken into account, shows that nations with endemic malaria grew around 1.3% less per person per year than non-endemic regions; however, a 10% reduction of malaria corresponds with an 0.3% increase in economic growth (Gallup & Sachs, 2001).

In addition to the economic impact of malaria on developing nations, there is of course an impact on first-world nations. While this impact is far less devastating for the latter, it is still worth a brief mention. Europe was declared free from malaria in the 1970's via a combination of drug use, environmental engineering (i.e. swamp draining) and the use of highly effective insecticides, and today most cases are imported by international travellers and immigrants (reviewed in Piperaki & Daikos, 2016). One study of malaria in the Slovak Republic covering 19 patients demonstrated that a preventative approach to the disease via chemoprophylaxis offsets costs incurred by direct medical intervention and time away from work, with savings of around 63.6% of what one would spend without chemoprophylaxis (Svihrova, *et al.*, 2012). Occupational exposure to malaria by 15 British soldiers stationed in Germany in 2001-2002 (malaria exposure was via a short peacekeeping mission in tropical malarious regions) led to a costings estimate

of €27 760 for their hospitalizations (Croft, *et al.*, 2005). In the United States of America, a country declared malaria-free in 1949, the estimated cost of treating a citizen who acquired malaria from abroad is around $25 250 USD (compared to prophylactic pre-travel treatment of $161 - $208 USD) (Adachi, *et al.*, 2014). Whilst overall these costs pale in comparison to the economic pains malaria gives regions such as Africa, South America and Asia, they are all costs which could be largely offset by a highly efficacious and affordable vaccine (which as a preventative would be more economically efficient than treatment, as demonstrated in the analysis by Stack, *et al.*, 2011). Therefore, wealthy nations should have an economic (in addition to a moral) interest in the eradication of malaria and a continued investment into vaccine development.

1.2.3 Anti-Malarial Vaccine Development

The history of anti-malarial vaccine development has been fraught with frustration and failure. Many a malariologist has had their dreams of saying "thank you" in Swedish ("tack så mycket") dashed. This section will briefly explore a modern historical look at the quest for the proverbial holy grail that is a safe, affordable and efficacious malaria vaccine. Due to space considerations and relevancy, mosquito-stage transmission-blocking and DNA-based vaccines will not be discussed (interested readers are referred to Hill, 2011 and Bergmann-Leitner & Leitner, 2013). For a general overview of some of the key developments towards said vaccine since the mid-20th century, see figure 1.2.

Figure 1.2. Key Developments in the Modern History of Malaria Vaccine Research: This flow chart focuses on some of the major immunological, molecular biochemistry, vaccinology and program developments since the 1940's. Abbreviations: WRAIR- Walter Reed Army Institute of Research, GSK- GlaxoSmithKline, WHO- World Health Organisation.

Modern malaria vaccine design has been borne out of the early research involving murine immunizations with parasite material (Cox, 1970), and has undergone various Kuhnian-style revolutions (Verhave, 2012) with respect to the best strategy in addition to a diversity of approaches used. Many of these revolutions have come about from novel understanding of immune responses against malaria, allowing (in conjunction with advances in technology such as recombinant protein production of vaccine components (Patarroyo, *et al.*, 1987)) for the development of a test vaccine that has always ultimately failed due in part to a paucity of more immunological knowledge and technological capability. Additional reasons for failure have also included a lack of scientific reason or rigour, such as the work by (Patarroyo, *et al.*, 1987). When these factors get updated by novel knowledge and technological progress, the cycle starts again.

One of the earliest approaches to malaria vaccine development has been the use of weakened sporozoite forms of the parasite (typically through irradiation); a whole-parasite approach. This grew out of the need to move away from killed blood-stage parasites, due to the requirements for new, effective and safe adjuvants to be used alongside them (reviewed in Mata, *et al.*, 2013). In the 1970's, humans inoculated with sporozoites from irradiated mosquitoes did demonstrate significant levels of protection, but only after a large number of mosquito bites (Clyde, *et al.*, 1975). The other problems with this approach were that not all parasites remained non-infective and protection waned after 3 months against *P. falciparum* (Clyde, *et al.*, 1975). More recently, biotech company Sanaria and collaborators not only revisited this approach, but improved on it with reports of 100% protection in 6 patients who were administered the vaccine known as PfSPZ (Seder, *et al.,* 2013). However, the problem with this approach was that five doses spread over the course of several months had to be used (with four doses, the protection fell to around 66% in nine patients where three developed malaria), and the vaccine had to be administered via the direct intravenous route (DIV); two issues that have practical problems for use of this vaccine for those most in need, i.e. very young children. An improved version of PfSPZ was developed, known as

PfSPZ-CVac, which consisted of *in vivo* chloroquine-attenuated sporozoites, that allowed dosing to be reduced to three lots in order to show 100% protection against malaria ten weeks after the third and final vaccination; though this was still via the DIV route (Mordmüller, *et al.*, 2017). As Adrian Hill mentions, additional challenges would of course be related to the immense costs involved in vaccine production and distribution due to the requirement for a cold-chain (Hill, 2011).

More advanced methods of generating sporozoite-based whole-cell vaccines have gathered some interest, such as using genetically-attenuated sporozoites that would not be able to break through to the blood-stage of infection, where merozoite morphological forms of the parasite cause disease. The first human study using genetically-attenuated parasites (GAPs) made use of mosquitoes infected with *P. falciparum* parasites with deletions of the *p52* and *p36* genes (important for liver stage development) (Spring, *et al.*, 2013). However, whilst all six-people remained blood-stage negative following low doses of the vaccine, high doses saw one person develop the presence of merozoites in peripheral blood. Despite this drawback, GAPs are still worth investigating and the use of further genetic deletions may prove fruitful.

An alternative version of the whole-parasite approach to malaria vaccine development is the use of chemically-attenuated blood-stage merozoites (reviewed in-depth in Stanisic & Good, 2015). Using the DNA-binding drug centanamycin to prevent parasite replication, *P. yoelii*-parasitized erythrocytes have been used to vaccinate mice, demonstrating nine months of protection against blood-stage infection (Raja, *et al.*, 2016). This data is of course early-stage, but somewhat encouraging; though the challenges of safely utilizing both DNA-binding drugs and *P. falciparum*-parasitized erythrocytes, in addition to regular immunological and efficacy studies remains.

The next major approach is of course the use of recombinant proteins for vaccine formulation. Currently, the most advanced is the RTS,s vaccine which is

comprised of a truncation of the circumsporozoite protein (CSP), the viral envelope protein of the hepatitis B virus and the AS01 chemical adjuvant; the latter two components required to increase the efficacy of the vaccine and immune responses (Stoute, *et al.*, 1997) (Kester, *et al.*, 2009). Phase III trials involved over 15 000 African children and after a year following the use of three doses of vaccine, yielded a 28% and 18% reduction in clinical malaria cases in young children and infants, respectively (RTS,S Clinical Trials Partnership, 2015) (Greenwood & Duombo, 2016). Those given a fourth dose demonstrated an average of 36.3% reduction in clinical malaria during 48 months of follow-up, which was estimated to have averted 1774 cases of clinical malaria per 1000 children (this number is due to the fact that multiple infections can occur). Whilst it has been recommended that the deployment of RTS,s go ahead (Greenwood & Duombo, 2016), concern should of course be given to the relatively poor level of longer-term protection.

Recently, a large study was published demonstrating the first use of a viral-vectored anti-malaria vaccine in a dose-escalation phase Ia study using 24 human volunteers (Payne, *et al.*, 2017). This study employed the use of the replication-deficient chimpanzee adenovirus serotype 63, and an attenuated orthopoxvirus-modified vaccinia virus Ankara (both established viral vectors in mammalian vaccine research), which encoded *pfrh5* from the 3D7 clone of *P. falciparum*. The vaccine was delivered via the intramuscular route in a heterologous prime-boost regimen utilizing an eight-week interval. All vaccinations were reported as being well-tolerated (with no serious adverse side-effects), and induced the production of anti-PfRh5 antibodies. These serum antibodies were used in a growth inhibition assay (GIA) to demonstrate an ablation of parasite invasion of erythrocytes, and were later demonstrated to target a mixture of linear and conformational epitopes on PfRh5. Interestingly, Payne *et al.* reported that the PfRh5-specific antibody responses were in excess of those seen in African adults with longer-term non-sterilizing immunity following years of exposure to malaria. What will become essential is to follow the immunological memory in each trial volunteer in order to

assess how long-lived the immune responses to the vaccine last. However, this study does, for the first time, show that a viral-vectored approach to malaria vaccine development is able to be legitimately explored in humans.

It should be argued that scientists must aim at developing a malaria vaccine that is multi-component and targets multiple stages of the parasite's lifecycle. This approach may help prevent the so-called "leakiness" of vaccines targeting the sporozoite (to prevent blood-stage breakthrough), decrease transmission of malaria (by targeting the sexual stages of the parasite), and ablate clinical malaria due to the blood-stage of parasite asexual development. Proteins that have been targeted as vaccine candidates include CSP, Pfs25, 48/45, and 230 (all of which are sporozite proteins), apical membrane antigen-1 (AMA-1), and PfRh5 (Cowman, *et al.*, 2016). Whilst these candidates should be pursued, proteins found to be essential to the blood stages of *P. falciparum* that are of immunological importance should also be further examined, such as PfRipr, which forms a necessary invasion complex with PfRh5 (Chen, *et al.*, 2011), (Volz, *et al.*, 2016). This will be further explored in section 1.4 - 1.6. However, first a succinct review of the parasite's life cycle will be given, so as to eventually narrow in on the clinically-important blood-stage of malaria.

1.3 Malaria Parasite Lifecycle Biology

Despite various geographical and phenotypic differences amongst malaria parasite species, they are all linked by a characteristically complex life cycle including a developmental stage taking place inside *Anopheles* mosquitoes. Refer to figure 1.3 for a summary.

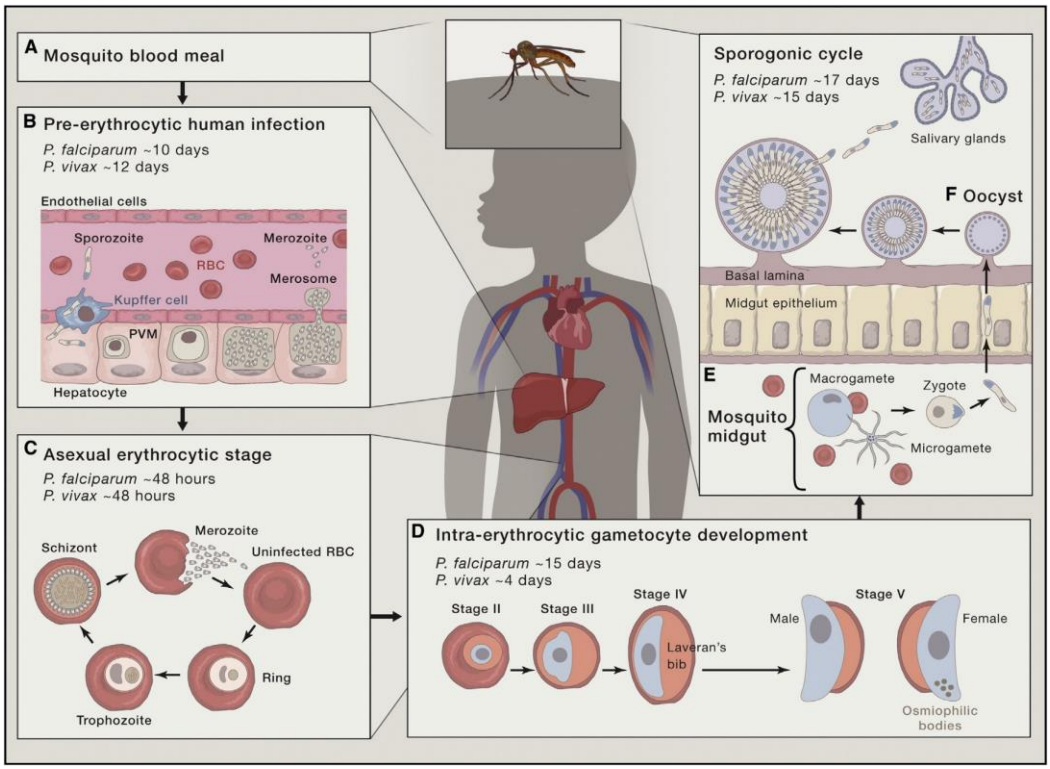


Figure 1.3. Lifecycle of the *Plasmodium* Parasite: A summary of the complex lifecycle of human malaria parasites. Figure from (Cowman, *et al.*, 2016). Following a mosquito blood meal (A), sporozoites move through the human skin to gain entrance to the blood stream in order to reach hepatocytes for their first round of asexual development (B). This leads to the emergence of merozoite morphological forms of the parasite that asexually reproduce inside erythrocytes (C). Sexual stages of the parasite can also emerge from alternative intra-erythrocytic development (D), which leads to male and female gametocytes that are taken up by feeding mosquitoes (E).

When a mosquito ingests parasitic gametocytes transmitted from a human host, they become activated in the insect's midgut lumen to produce gametes in a process that depends on various factors including pH, a temperature drop, increased levels of calcium and the presence of xanthurenic acid (Guttery, *et al.*, 2012). The resulting zygote occurring from the fusion between a male and female gamete develops into a highly motile ookinete that penetrates the mosquito's gut wall epithelium, facilitating differentiation into an immobilized oocyst (Guttery, *et al.*, 2012) (Doi, *et al.*, 2010). This oocyst becomes the developmental site of new

sporozoites that, upon emergence, migrate to mosquito salivary glands where they wait to be injected into the next human host when the mosquito's penetrating proboscis strikes host dermal tissue (Doi, *et al.*, 2010).

Up to approximately 120 motile sporozoites will find themselves in the host's dermis, where the parasites spend up to three-hours before trickling into the blood stream, and occasionally the lymphatics (Tamauchi, *et al.*, 2007). Once blood-borne, sporozoites head to the liver, crossing the sinusoidal wall via gliding motility, in a process taking as little as two-minutes (Doi, *et al.*, 2010). Sporozoites traverse various cell types such as Kupffer cells and fenestrated endothelial cells before finally entering a hepatocyte to undergo merogony (Tavares, *et al.*, 2013). One of the key signals for hepatocyte invasion to occur is the parasitic recognition of highly sulphated forms of heparin sulphate proteoglycans (HSPGs) which triggers calcium-dependant protein kinase 6 and the binding of CSP to HSPG and subsequently, AMA-1 adhesion (Coppi, *et al.*, 2007) (Herrera, *et al.*, 2015). Once inside a hepatocyte, the parasite sets up a parasitophorous vacuole via interactions between its p52 and p36 proteins and the hepatocyte receptor EphA2; essential molecular events prior to the subsequent two – ten days of merogony (Kaushansky, *et al.*, 2015). It should be noted that *P. vivax* and *P. ovale* parasites can become hypnozoites to establish latent infection (reviewed in detail in (Prudêncio, *et al.*, 2006)).

The liver stage of infection is responsible for the asexual production of approximately 40 000 merozoites per hepatocyte, which are released into the blood following their release from merosomal vesicles (Sturm, *et al.*, 2006). The resulting blood-stage of infection is where the symptoms of clinical malaria manifest as parasites undergo rapid asexual multiplication inside erythrocytes via a repeating sequential process of pre-invasion set-up, active invasion, echinocytosis and host cell rupture (Weiss, *et al.*, 2015). The initial steps of blood-stage infection are given in figure 1.4.

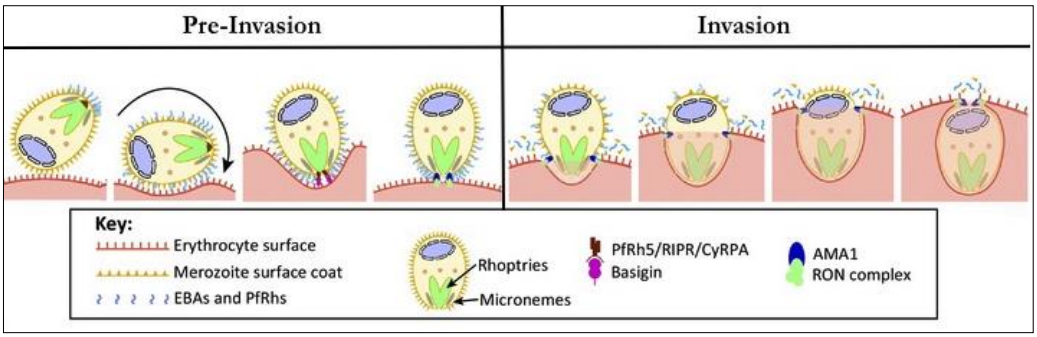


Figure 1.4. Merozoite Invasion of Erythrocytes: Pre-invasion is a process taking around ten-seconds, and involves a host of molecular interactions between the merozoite surface coat proteins and the erythrocyte surface leading to its deformation. Reorientation occurs so that the apical tip of the merozoite is perpendicular to the erythrocyte, with EBA and PfRh proteins (originally stored in micronemes and rhoptries) of the former engaging with ligands of the latter. Once PfRh5 of the PfRh5-associated invasion complex (PAIN-complex) engages BSG and the AMA-1/RON complex is established, the parasite has committed to invasion. Invasion lasts again approximately ten-seconds and involves the enzymatic shedding of merozoite surface proteins whilst the tight moving junction created by the AMA-1/RON complex in addition to reverse mobilization of parasitic actomyosin allows the parasite to enter the erythrocyte. During entry, a parasitophorous vacuole is established. Figure and text adapted from (Weiss, *et al.*, 2016).

The glycophosphatidylinositol (GPI) anchored protein, merozoite-surface protein-1 (MSP-1), is the predominant surface coat protein the merozoite has been hypothesized to use for initial attachment to erythrocytes (Holder, 1994). However, evidence from genetic knockdown studies of MSP-1 reveal it has a level of redundancy whereby merozoites are still capable of engaging with and invading erythrocytes (Das, *et al.*, 2015). It is possible that aside from MSP-1, other epidermal growth factor-like module-rich proteins (EGFs) amongst the merozoite surface coat are important for the initial attachment events such as MSP2, MSP4, Pf's 12, 38 and 92 (potential candidates are extensively reviewed in Crabb, 2006 and Beeson, *et al.*, 2016). Nonetheless, after initial attachment, the erythrocyte is physically deformed at the point of initial parasitic contact, which is followed by a reorientation event resulting in the merozoite's apical tip lying perpendicular to the erythrocyte surface. This reorientation event was until recently (Bargieri, *et al.*,

2016) thought to be, at least in part, due to the merozoite thrombospondin-related anomalous protein (MTRAP), which binds aldolase (an actin-binding protein) through its cytoplasmic domain (Baum, *et al.*, 2006). MTRAP was also believed to play a vital role in tight junction formation (Baum, *et al.*, 2006), though is not utilized directly for erythrocytic invasion itself (Riglar, *et al.*, 2016). It is now understood that both of these claims are inaccurate, as CRISPR/Cas9-based genetic disruption work on blood stages revealed no observable defects in parasitic growth recently (Bargieri, *et al.*, 2016), so clearly more investigation around the reorientation event during invasion is warranted.

Next in the pre-invasion phase of invasion is the binding between apical tip proteins and key erythrocyte receptors; important with respect to the former are the erythrocyte-binding antigens (EBA's) and the *P. falciparum* reticulocyte binding-like homologs (PfRh's). EBA proteins bind to host glycophorins and the PfRh family of proteins including PfRh4 and PfRh5 engage a variety of receptors such as complement receptor 1 (CR1) and BSG, respectively (the PfRh family of receptors will be looked at in more detail in section 1.4, those interested in the EBA family are referred to Tham, *et al.*, 2012, Cowman, *et al.*, 2016, and Wassmer & Carlton, 2016). Significant erythrocyte deformation occurs following the binding between EBA-175 and glycophorin A, resulting in a phosphorylation cascade and subsequent alterations in host cell cytoskeletal proteins (Sisquella, *et al.*, 2017) (Koch, *et al.*, 2017). In addition to the interactions described, Ca^{2+} signalling inside the invading merozoite occurs, with phosphorylation of the cytoplasmic tail of PfRh4 taking place during its engagement to CR1 (Tham, *et al.*, 2015). At around the same time as host cell deformation is occurring, the erythrocytic BSG gets bound to by PfRh5 in the PfRh5-Associated INvasion-complex (PAIN-complex), consisting of the cysteine-rich protective antigen (CyRPA) (Reddy, *et al.*, 2015) and the *P. falciparum* Rh5-Interacting PartneR (PfRipr) (Chen, *et al.*, 2011). This step is essential for invasion to occur, with potent inhibition of invasion resulting from experimental models using *bsg* knockouts, soluble recombinant BSG or complete invasion inhibition via anti-BSG antibodies (Crosnier, *et al.*, 2011).

Following the resultant Ca^{2+} influx into the erythrocyte due to PAIN-complex binding (Volz, *et al.*, 2016), the merozoite forms an irreversible tight junction once RON2 is inserted into the host cell membrane and binds to AMA-1 found on the parasitic cell surface, thus committing the parasite to invasion (Besteiro, *et al.*, 2011). This results in the formation of the RON/AMA-1 complex and secretion from rhoptries of lipid-rich contents (likely involved in the formation of the parasitophorous vacuole), merozoite surface coat protein shedding via enzymatic cleavage, and activation of the parasitic actomyosin motor (Riglar, *et al.*, 2011). Upon entry into the erythrocyte, the parasitophorous vacuole is sealed around the parasite, protecting it from the local environment and serving as one of the barriers upon which parasitic proteins are exported towards the host cell surface via the *P. falciparum* translocon of exported proteins (PTEX) (Koning-Ward, *et al.*, 2009). The infected erythrocyte experiences echinocytosis a process defined morphologically by cell-shrinkage and formation of spikey protrusions, which may be due to the Ca^{2+}-influx that follows PAIN-complex/BSG binding (Weiss, *et al.*, 2015).

Once invasion has occurred, the parasite begins extensive remodelling of the erythrocyte, essentially setting up a temporary "home" from which it can evade the host immune system and undergo asexual reproduction in 48-hour cycles leading to the development of up to 32 additional merozoites per infected cell (the intra-erythrocytic stage of development is reviewed extensively in Boddey & Cowman, 2013). Inside the infected erythrocyte, an extensive protein trafficking network is established (see figure 1.5), with the *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) being the most important parasitic protein set up on the erythrocyte cell surface, subsequently allowing for sequestration onto capillary beds thus evading splenic clearance (Su, *et al.*, 1995) (Cranston, *et al.*, 1984). Furthermore, parasitic proteins that contribute to host cell remodelling allow the infected erythrocytes to withstand higher temperatures that develop due to malarious fever, in addition to

being able to tolerate higher shearing forces due to increased rigidity via knob-like protein complexes (Maier, *et al.*, 2008) (Crabb, *et al.*, 1997).

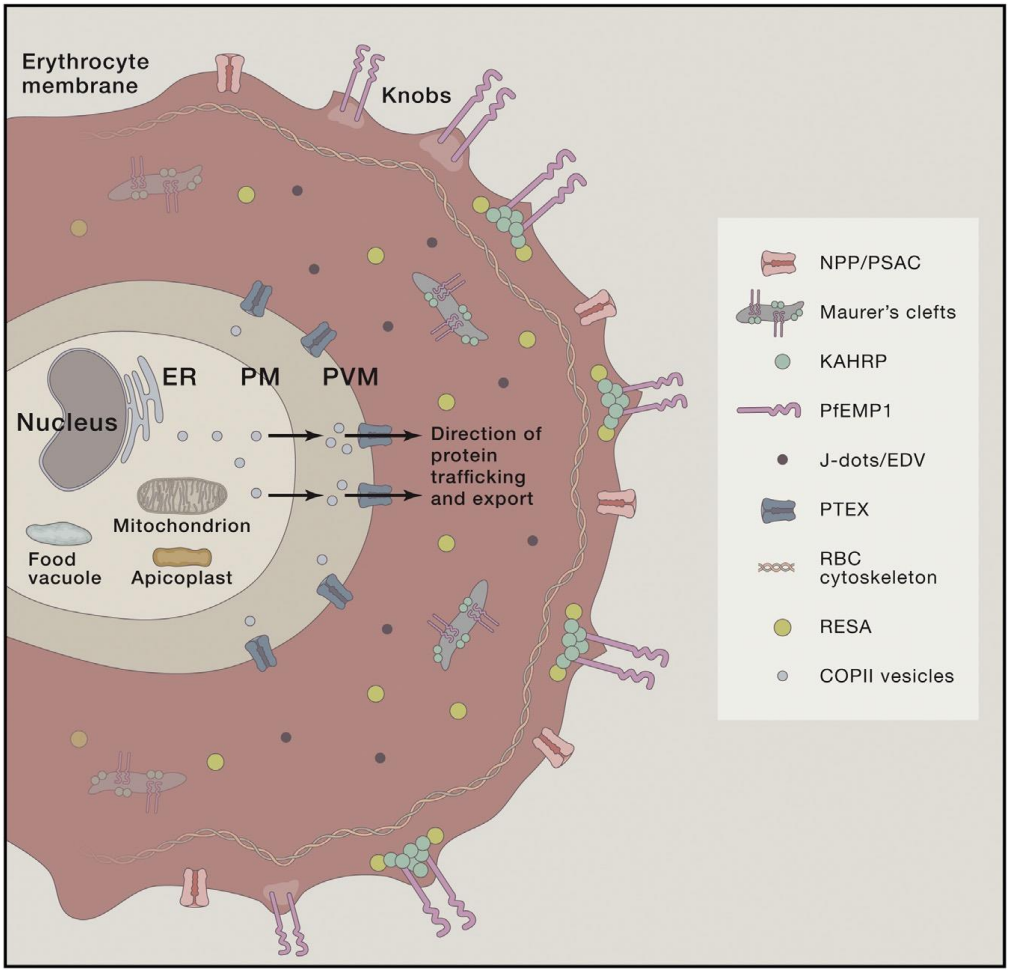


Figure 1.5. Architecture of the Infected Erythrocyte: This erythrocytic cross-section provides a glimpse of the extent of remodelling of the infected host cell. Within the parasite membrane (PM) lies various *Plasmodium* organelles including the endoplasmic reticulum (ER) that facilitates the vesicle-mediated transport of proteins to the erythrocyte cytoplasm and cell surface, via PTEX on the parasitophorous vacuole (PVM). The food vacuole is responsible for the metabolism of haem, which is too toxic to malaria parasites, and so is converted into haemozoin. Abbreviations not elsewhere cited: NPP/PSC- new permeability pathways, KAHRP- knob-associated histidine-rich protein, EDV- electron-dense vesicles, RBC- red blood cell, RESA- ring-exported surface antigen, COP- coat protein. Image and information from (Cowman, *et al.*, 2016).

High parasitaemia can develop during the asexual reproduction (schizogony) of *P. falciparum* parasites in the blood due to the exponential development of around 10-fold parasite populations per 48-hour cycle that facilitates a cytokine-mediated inflammatory response correlating with an increase in disease severity (and death) (Kwiatkowski, *et al.*, 1990). During schizogony a proportion of parasites switch to producing sexual forms termed gametocytes. This switching results from key microenvironment changes such as drug exposure (not limited just to antimalarial drugs, but also steroid hormones), dense parasitaemia levels, parasitic endoplasmic reticulum stress, anaemia (especially sickle-cell anaemia) and erythrocyte haemolysis (reviewed extensively in (Ngwa, *et al.*, 2016)). The parasites inside infected erythrocytes are capable of signalling each other to allow for an increase in gametocytogenesis via extracellular vesicles containing DNA, RNA and proteins (Regev-Rudzki, *et al.*, 2013). The egress of these sexual forms of the parasite from erythrocytes requires MTRAP in order to lyse the parasitophorous vacuole membrane (Bargieri, *et al.*, 2016). Once gametocytes are produced and taken up by a feeding mosquito, the entire cycle begins again.

1.4 An Examination of Key Apical Merozoite Proteins

As demonstrated previously in figure 1.4 and section 1.3, the EBA and PfRh proteins of the merozoite play an important role in binding erythrocyte receptors for invasion (figure 1.6), and therefore to this key step in the lifecycle of *Plasmodium* parasites. EBA proteins are secreted from the micronemal organelles, with EBA-140, 175 and 180 possessing conserved Duffy binding-like domains (referred to as F1 and F2) that bind erythrocyte receptors (Adams, *et al.*, 1992) (Adams, *et al.*, 2001). Erythrocyte glycophorins A, B and C bind to EBA-175, erythrocyte binding ligand-1 (EBL-1) and EBA-140, respectively, with EBA-181 potentially interacting with the erythrocyte membrane protein band 4.1; though the definitive binding partner remains elusive (Sim, *et al.*, 1994) (Mayer, *et al.*, 2008) (Lobo, *et al.*, 2003) (Gilberger, *et al.*, 2003) (Lanzillotti & Coetzer, 2006).

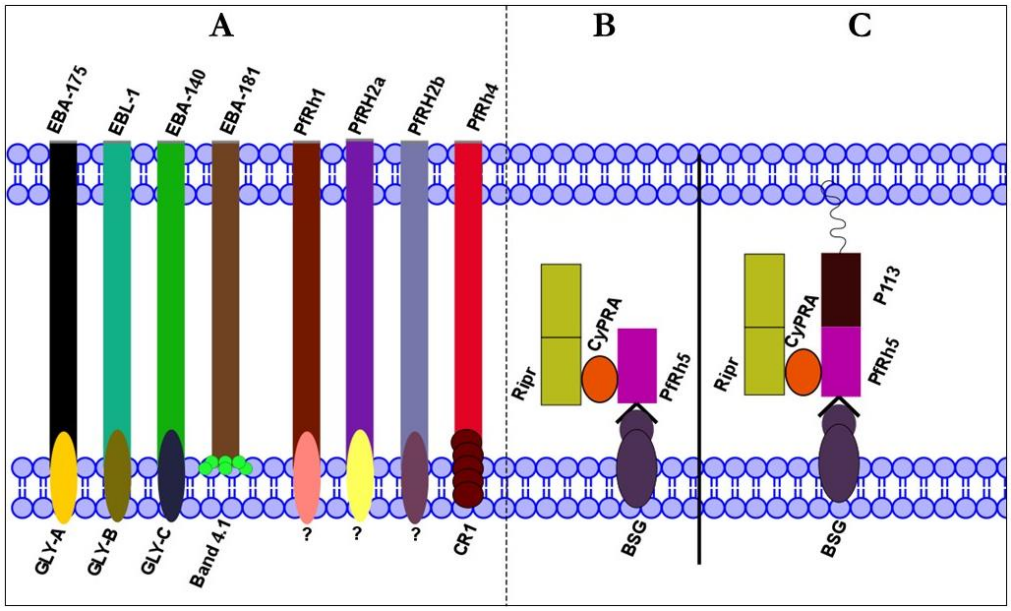


Figure 1.6. Key *P. falciparum* Proteins and their Association with Erythrocytic Ligands: A. This figure demonstrates the EBA and PfRh family of proteins found on the apical tip of the parasite and their cognate erythrocytic binding partners (for those known; the erythrocytic binding partners for PfRh1 and 2a/b are as yet unknown, and the binding partner for EBA-181 is assumed to be Band 4.1). B. As demonstrated, host cell BSG is engaged by the proteins of the PAIN-complex on the apical tip of the merozoite, with direct association due to PfRh5. B. One of two current models (the other being C) whereby the interaction between the PAIN-complex and the merozoite membrane is unknown. C. A recent model developed showing that the PAIN-complex is tethered to the merozoite membrane via P113, which also binds PfRh5. Figure made by Mr. William Godfrey (with guidance from the author).

Aside from PfRh3 (a transcribed pseudogene), modelling studies based on the solved crystal structure of PfRh5 suggest that PfRh's 1, 2, and 4 share its kite-like structure despite low sequence homology (Wright, *et al.*, 2014) (Chen, *et al.*, 2014). PfRh5 is unique amongst the PfRh proteins in that it is significantly smaller at around 65 kDa compared to >200 kDa for the others, and is refractory to genetic deletion in all strains tested (Baum, *et al.*, 2009). PfRh1 functions prior to the other proteins, and once bound to the as yet unknown host cell receptor, allows for a rise in merozoite Ca^{2+} levels and subsequent release of EBA-175 from micronemes (Gao, *et al.*, 2013). PfRh2a and PfRh2b proteins are mostly identical with respect to their N-terminal regions, with significant divergence occurring from 42 and 54

kDa, respectively, from the cytoplasmic C-terminus (Jayner, *et al.*, 2001) (Triglia, *et al.*, 2011). PfRh4 binds to CR1, specifically the complement control protein (CCP) modules 1-3 (Tham, *et al.*, 2010) (Tham, *et al.*, 2011), and is upregulated when there is a loss of EBA-175 function (Stubbs, *et al.*, 2005). Due to the focus of this thesis, a closer look at PfRh4 and CR1 will be given, followed by an examination of BSG and the PAIN-complex proteins.

1.4.1 A Closer Examination of PfRh4 and Complement Receptor 1

In vitro studies of *P. falciparum* blood culture reveal two major erythrocyte invasion pathways; sialic acid (SA)-dependent and SA-independent. A variety of erythrocytic receptors utilized by *Plasmodium* parasites possess SA moieties which serve to confer a net negative charge to erythrocytes (for example, the glycophorins, band 4.1 and BSG); the exception to this is the sialic acid-deficient CR1 protein (Tham, *et al.*, 2012). Treating erythrocytes with neuraminidase (NA) to catalyze the hydrolysis of SA residues present has been shown to compromise the merozoite invasion process (Baum, *et al.*, 2003). However, certain strains of *P. falciparum*, such as W2mef parasites, have demonstrated the ability to switch from SA-dependent to SA-independent invasion via continued culturing in NA-treated human blood cultures or disruption of EBA-175 (Stubbs, *et al.*, 2005) (Dolan, *et al.*, 1990). PfRh4 was found to be responsible for the molecular basis of this switching (Stubbs, *et al.*, 2005). This was discovered after initially demonstrating that invasion pathway switching was reversible (from SA-independent to SA-dependent and back again) with clonal lines of W2mef parasites, and then using microarray and RT-PCR analysis a 60-80-fold increase of PfRh4 transcription in SA-dependent lines compared to SA-independent parasite lines. Next *PfRh4* disrupted parasites (W2mefΔRh4) were constructed, and from NA-treated blood cultures these parasites as well as a variety of SA-dependent *P. falciparum* lines showed no detection of PfRh4 with PfRh4-specific antibodies, in contrast to SA-independent strains. Finally, it was revealed that the W2mefΔRh4 parasites could not switch to

SA-independent invasion, and that CSL2 clonal lines of *P. falciparum* parasites silence active *PfRh4* over time when re-cultured in regular erythrocyte cultures. What would be an ideal experiment to still do is to develop a knock-out/knock-in model whereby *pfrh4* was able to be disrupted, cultured as described above by Stubbs *et al.,* then restore functionality of PfRh4 to see if the parasites were able to grow again on SA-containing red cell cultures. This would conclusively prove the necessity for PfRh4 in SA-dependent invasion of erythrocytes.

Of course, erythrocytes treated with neuraminidase occurs only with *in vitro* conditions rather than what is seen in the field; however, this is analogous to situations where the host immune response may prevent certain invasion pathways, or polymorphisms affecting erythrocytic surface proteins which affect the choice of receptors for the parasite (Tham, *et al.*, 2010). Three types of predominant polymorphisms occur for CR1 (as reviewed by Krych-Goldberg & and Atkinson, 2001); firstly, the number of complement control protein repeats (CCPs) and their amino acid number in each of the four allelic variants of CR1, second is the carriage of Knops blood group antigens such as Swain-Langley (Sl or Sl^{a}) on CR1, and thirdly is the copy number of CR1 on erythrocytes. Of note, Sl^{a-} individuals may be afforded some level of protection against severe malaria caused by *P. falciparum* erythrocyte membrane protein 1 (PfEMP1)-mediated resetting of erythrocytes (Rowe, *et al.*, 1997). Additionally, the presence of the Sl^{a-} phenotype is approximately 70% in West Africans, 40% in African-Americans and only around 1% in Caucasians, giving potential evidence in support of the hypothesis that the Sl^{a-} phenotype has been selected for in malarious regions (Moulds, *et al.*, 2000). However, in 2012, research emerged demonstrating a lack of evidence that the Knops polymorphisms occurring on the fourth long homologous repeat of the CR1 protein are indeed driven by malaria, due to a lack of structural or functional modulation of CR1 (Tetteh-Quarcoo, *et al.*, 2012). With respect to copy number, certain Papua New Guinean populations in malaria-burdened areas possessed around 200 CR1 molecules per erythrocyte (the lowest level in the world), which helped confer resistance to severe malaria (Cockburn, *et al.*, 2004).

As previously mentioned, no structure of PfRh4 exists. PlasmoDB (www.plasmodb.org) lists PfRh4 as having a molecular weight of 205 kDa; however, following signal sequence cleavage and proteolytic processing events, a 180-190 kDa size can be seen on Western blots following the use of samples from saponin-treated schizont pellets, and 160 kDa from culture supernatants (Tham, *et al.*, 2009). Binding to CR1 occurs through a region in PfRh4 located towards the N-terminus, potentially within a 58-amino acid region located between residues 282 and 340 (Tham, *et al.*, 2009). Further studies are required to further narrow down the key residues involved in binding to CR1, in addition to the 3-dimensional structure of PfRh4.

In contrast to PfRh4, a lot more information is known about CR1, most likely due to its well-known role in the immune system as a regulator of complement, and because it carries the polymorphic Knops blood group antigens (Moulds, *et al.*, 2001) (Moulds, *et al.*, 2000), which are of importance in immunohaematology. In humans, CR1 is classified as a type 1 transmembrane glycoprotein, with the extracellular region existing as a single chain comprised of independently folding CCP domains (see figure 1.7A) (Krych-Goldberg & and Atkinson, 2001). Starting from the N-terminus, groupings of seven CCP's at a time comprise four long homologous repeats (LHR's; A-D) in the most common size variant of CR1. Erythrocytic CR1 binds opsonized immune complexes for transport to the spleen and liver, with C4b binding to CCP1-3 (site 1), and C3b/C4b binding to CCP8-10 and 15-17 (site 2). Note that this activity of site 1 and 2 are enhanced if key amino acids are mutated to confer a positive charge or remove a negative charge at key residues (Krych, *et al.*, 1998).

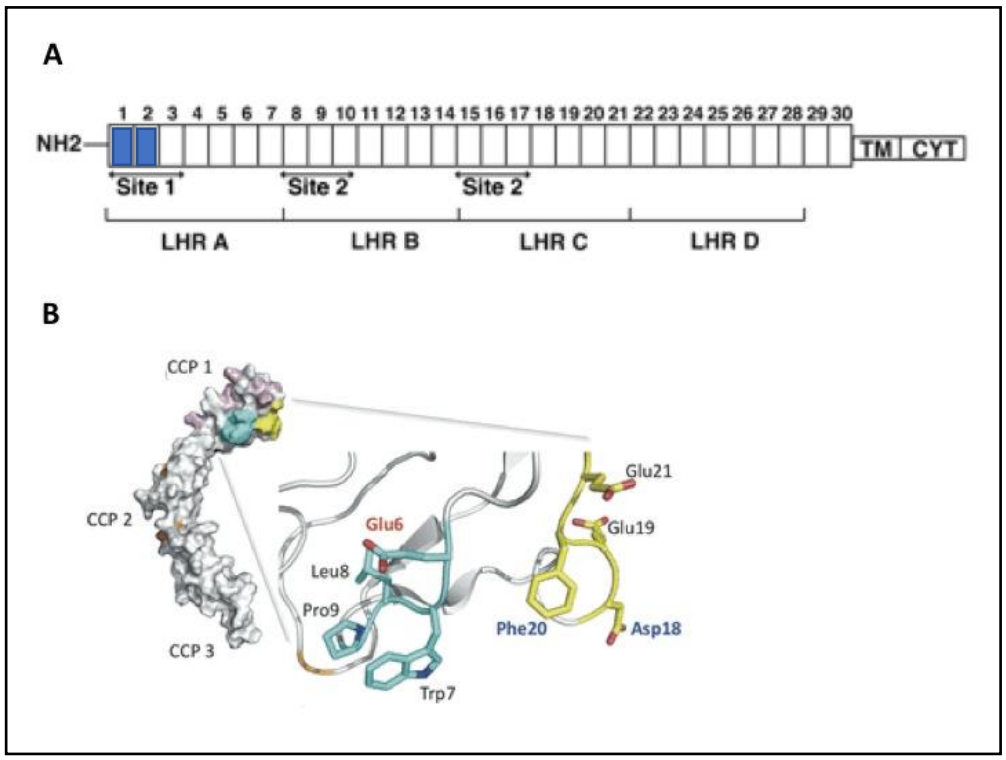


Figure 1.7. Structural Information Regarding CR1: A. Schematic of the overall architecture of CR1, which is comprised of 30 CCP modules. Site 1 is responsible for decay accelerating activity (with the CCPs responsible for binding to CR1 coloured blue), and site 2 CCP's are required for co-factor activity allowing for cleavage of C3b and C4b complement proteins. B. NMR-derived structure of CCP1-3, including key residues involved in binding PfRh4. Figure adapted from (Krych-Goldberg and Atkinson, 2001) and (Park, *et al.*, 2013). Abbreviations used: TM, transmembrane domain; CYT, cytoplasmic tail; LHR, long homologous repeats.

Aside from erythrocyte CR1 possessing domains crucial to the functioning of the immune system, there are regions within the LHR's exploited by *P. falciparum* for both the development of blood-stage pathology and parasite invasion of erythrocytes. LHR's B and C are required for the formation parasite-infected erythrocyte rosettes, with Rowe *et al.* mapping the binding site of PfEMP1 (the parasite protein expressed on the surface of infected erythrocytes) to the C3b-binding regions that make up site 2 of CR1 (Rowe, *et al.*, 2000). However, with respect to PfRh4, site 1 of LHR-A had been identified as the binding site for this parasite protein, used to assist merozoite invasion of erythrocytes (Tham, *et al.*,

2011). Additionally, evidence for the formation of a ternary complex between PfRh4, C4b and site 1 of CR1 had been found, with no apparent effect on the affinity for C4b binding, following the complex formed by PfRh4 and CR1 (Tham, *et al.*, 2011). Glutamic acids 6, 19 and 21, tryptophan 7, aspartic acid 18 and phenylalanine 20 (surface exposed) have all been identified as key residues in CCP1-3's binding site for PfRh4, with the latter two amino acids being absolutely essential (see figure 1.7B) (Park, *et al.*, 2013).

1.4.2 A Closer Examination of Basigin with Respect to PfRh5

Basigin (BSG) is an antigen of the Ok blood group (defined by an allogenic difference of BSG), and is defined as a transmembrane glycoprotein of the immunoglobulin superfamily (Muramatsu, 2012). Four isoforms exist due to alternative splicing and alternative transcription initiation, with BSG-2 being the most common isoform (Muramatsu, 2012). Both BSG-1 and 2 interact with PfRh5; however, it is believed that the shorter BSG-2 is the isoform more commonly expressed on erythrocytes (Crosnier, *et al.*, 2011). For some time, there had been some disagreement as to whether the interaction between BSG and PfRh5 is NA-resistant or sensitive. This was based on published results indicating said NA-sensitivity (Hayton, *et al.*, 2008). However, Baum *et al.* presented results suggestive of binding occurring in a SA-independent, NA-insensitive manner (Baum, *et al.*, 2009) (which complemented independent research demonstrating that the aforementioned interaction is refractory to NA treatment (Rodriguez, *et al.*, 2008)). Additionally, the fact that recombinant PfRh5 binds to recombinant non-glycosylated BSG proves the interaction is NA-insensitive (Crosnier, *et al.*, 2011). It should also be noted that it is the specificity of BSG for PfRh5 that is predominantly responsible for the host tropism of *P. falciparum,* with chimpanzee BSG binding of much lower affinity and no detectable binding of gorilla BSG, as shown by AVEXIS and SPR experiments (Wanaguru, *et al.*, 2013). Of note, the

sequence similarity between human and either chimpanzee or gorilla BSG is approximately 99.03% and 97.09% homologous (Forni, *et al.,* 2015).

As BSG has four isoforms and is highly glycosylated, its molecular weight ranges from 43 – 66 kDa (Muramatsu, 2012). The crystal structure of the extracellular region was solved in 2008, demonstrating a protein rich in β-sheets (figure1.8) (Yu, *et al.*, 2008). The transmembrane region possesses a 23-amino acid segment conserved between human, mouse and chicken BSG (Muramatsu, 2012). Furthermore, the extracellular domain of BSG-2 contains three possible asparagine glycosylation sites (these are regions of differential glycosylation), which may help the extracellular portion of the protein stand upright, allowing maximum exposure of ligand binding regions (Yu, *et al.*, 2008). BSG is a unique protein with a diverse array of functions including membrane-bound association with monocarboxylic acid transporters (MCT) for cellular metabolism (actually bound to a BSG dimer), integrin binding within the same membrane (including a super-complex of BSG dimers, MCT's and β_1-integrin or CD44), cyclophillin A binding via BSG residues 207-214 of the transmembrane region for regulation of chemotaxis, and association with platelet glycoprotein VI (Muramatsu, 2012).

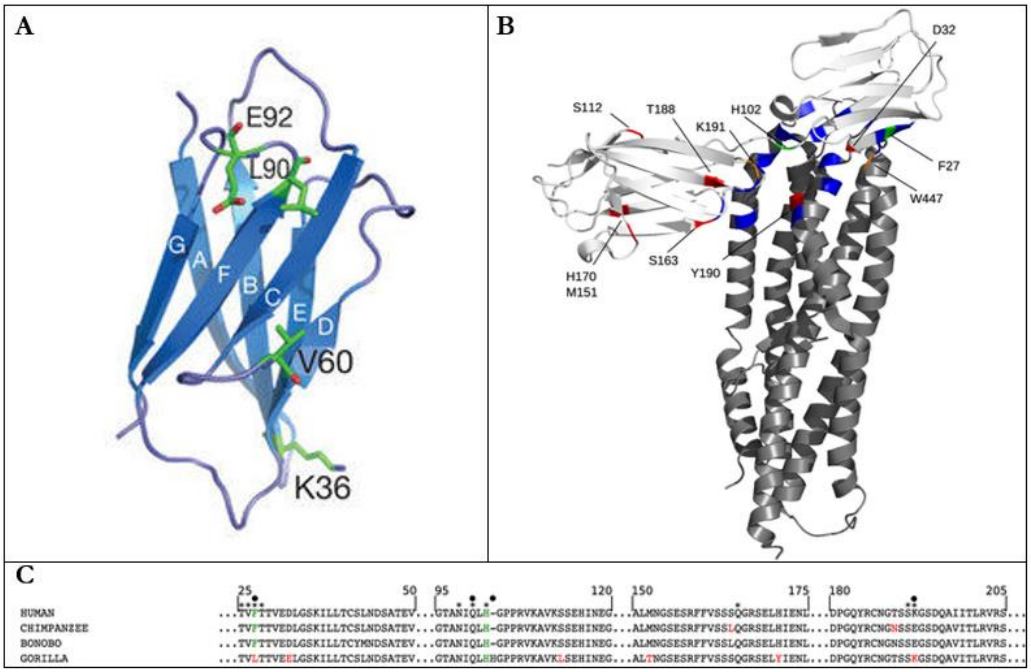


Figure 1.8. Structure of Basigin: A. The structure shown is of the membrane-distal immunoglobulin superfamily region of BSG-2. β-sheets are lettered from A-H, and the amino acids listed are locations of naturally-occurring variants. Figure adapted from Crosnier, *et al.*, 2011. B. Human BSG (shown in white) bound to PfRh5 (grey). Residue W447 from PfRh5, in addition to BSG positively selected sites that interact with PfRh5, are shown in orange; residues involved in the interaction are shown in blue. C. Multiple primary sequence alignment of BSG for human and a few representative primate species are shown. B and C are adapted from Forni, *et al.,* 2015.

It should be mentioned that the Ok blood group is the result of a polymorphism within BSG and that when the glutamic acid residue at position 92 is changed to a lysine, the Ok^{a-} blood group results (Spring, *et al.*, 1997). The Ok^{a-} blood group version of BSG has less affinity for PfRh5 and are more resistant to *P. falciparum* infection (Crosnier, *et al.*, 2011). Furthermore, it may be important to note that the glutamic acid 92 is located just above the glycan-free β-sheets of the membrane-distal domain of BSG (as shown in figure 1.8 A), making steric hindrance by glycan moieties less likely (Muramatsu, 2012). Additionally, the structure of human BSG plays an important role for the species-specific tropism of *P. falciparum* infection. In 2013, Wanaguru *et al.* identified a lysine amino acid at position 191 in human erythrocytic BSG to be important for PfRh5 affinity, and the presence of

an oppositely charged glutamic acid residue on chimpanzee erythrocytic BSG contributing to far binding lower affinity, and no detection of gorilla BSG binding to PfRh5 due to multiple residue differences (figure 1.8 B and C) (Wanaguru, *et al.*, 2013). This work confirmed previous results (Crosnier, *et al.*, 2011) that both immunoglobulin-like domains of BSG are required to recognize and bind PfRh5.

1.4.3 A Closer Examination of the PAIN-complex Proteins

PfRh5 is a 63 kDa protein, making it far smaller than its other PfRh family members; however, a processed 45 kDa fragment, which appears during invasion demonstrates greater stability and shows no expression level variation across strains (Rodriguez, *et al.*, 2008) (Chen, *et al.*, 2011) (Baum, *et al.*, 2009). The cleavage site for processing the 45 kDa protein exists between residues 125 and 135 of the 63 kDa protein (Chen, *et al.*, 2011). This fragment has also been demonstrated to bind to erythrocytes with greater affinity than full-length PfRh5 (PfRh5FL) and was therefore designated to have an important function (Wright, *et al.*, 2014). It should also be mentioned that PfRh5 is highly conserved among *P. falciparum* parasites, with a mere five common non-synonymous single nucleotide polymorphisms (SNPs) (Williams, *et al.*, 2012) (Bustamante, *et al.*, 2013) (Reddy, *et al.*, 2014). Importantly, antibodies against one recombinant PfRh5 variant were capable of neutralizing *P. falciparum* parasites of all heterologous strains tested, which contained the aforementioned SNPs (Douglas, *et al.,* 2011) (Williams, *et al.*, 2012) (Bustamante, *et al.*, 2013). Additionally, natural-acquired anti-PfRh5 antibodies from the sera of immune individuals are capable of preventing parasite growth *in vitro* via GIA's and were predictive of protection against severe malaria (Tran, *et al.,* 2013).

In 2014, two crystal structures of PfRh5 were published; one a 48 kDa fragment (Chen, *et al.*, 2014), the other a 45 kDa fragment bound to BSG (Wright, *et al.*, 2014) (figure 1.9). The structures revealed an overall ellipsoid, kite-like structure

including a core, formed of nine largely antiparallel α-helices encasing a β-hairpin towards the N-terminus. Outermost helices were shown to contain considerable kinks and breaks. The secondary structure of the molecule includes a triplet helical coiled-coil domain orientated along the long axis of the ellipsoid, in addition to a short three-helix bundle plus short two-helix coiled-coil domains. The domains were revealed to be loosely packed within the structure. Two disulfide bonds were seen in the structure, one of which holds the two halves of the structure together (C345 – C351), the other linking two helices together near a kink (C224 – C317) (Chen, *et al.*, 2014). The disulfides were shown to be important to the stability of the protein rather than being directly involved in protein interactions. PfRh5FL was shown to have six cysteine residues in total in certain parasite lines but only five in others, which indicates that cysteine residues may be unpaired or buried within the protein (Wright, *et al.*, 2014).

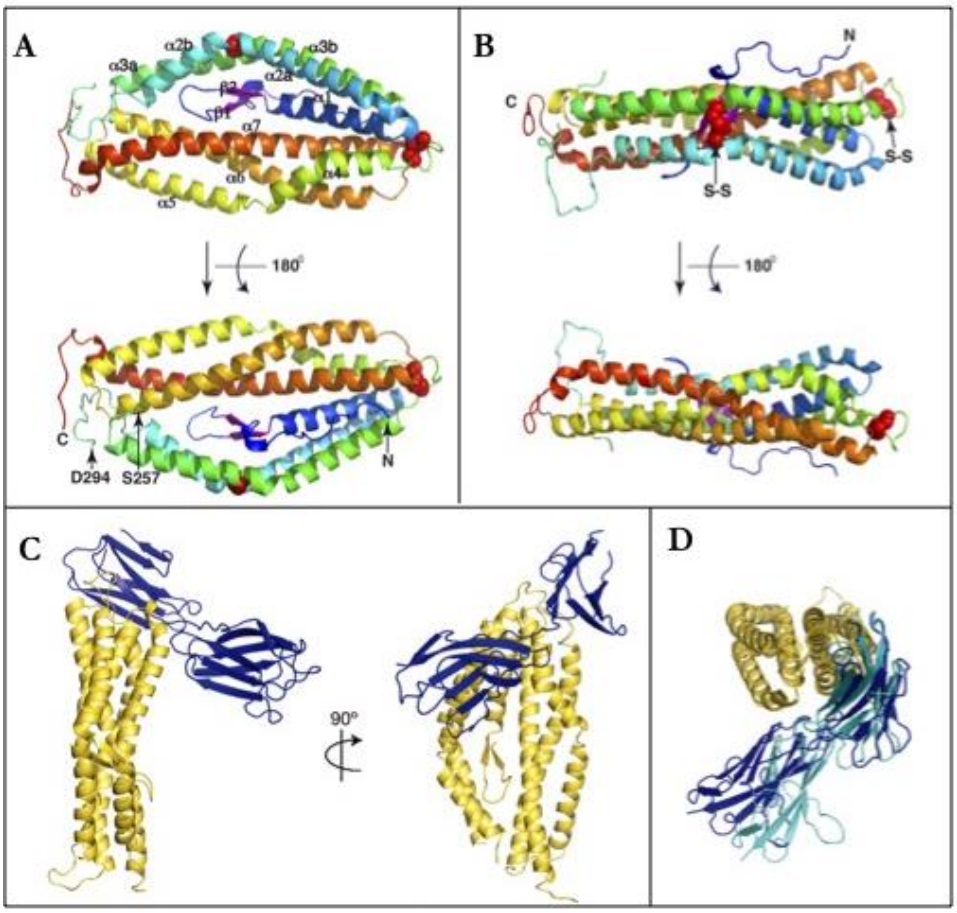


Figure 1.9. Crystal Structures of PfRh5: A. Ribbon-style representation of PfRh5 with N-terminus in blue and C-terminus in red, with each α-helix labelled α1-7. The β-hairpin is shown in magenta and red spheres indicate cysteines forming disulfide bonds (C345 - C351, and C224 – C317). The residues from D294 and S257 are shown to be disordered. B. PfRh5 structure viewed from the side. C. PfRh5 in yellow, shown bound the BSG. D. Top-down view of BSG bound to PfRh5 displaying both conformations that were displayed in an asymmetric unit (blue and cyan). Figures A and B were solved to a resolution of 2.18 Å from X-ray diffraction data (figure adapted from Chen, *et al.*, 2014). Figures C and D were from X-ray diffraction data solved at a resolution of 3.1 Å (figures adapted from Wright, *et al.*, 2014).

PfRh5 lacks a transmembrane domain and cytoplasmic tail, allowing it to be deduced that it forms a complex with another protein/s in order for anchoring to the merozoite membrane. Previous speculations proposed PfRipr as the anchor (Baum, *et al.*, 2009), but later avidity-based extracellular interaction screening (AVEXIS) assays appeared to show 19 amino acids of the N-terminal region of PfRh5

interacting with the merozoite surface protein P113 (Galaway, *et al.*, 2017). P113 (also known as Pf113) was initially discovered as a highly abundant protein in merozoite high molecular mass complexes during studies aimed at identifying detergent-resistant membrane proteins (Sanders, *et al.*, 2007) (Sanders, *et al.*, 2005) (Gilson, *et al.*, 2008). The P113 protein is reported to contain a glycosylphosphatidylinositol (GPI) anchor, and appeared to be surface localized via antibody markers resulting in staining both on non-permeabilized and permeabilized merozoites (Gilson, *et al.*, 2008).

The recent AVEXIS and SPR assays appeared to show that P113 can interact with the N-terminus of PfRh5 (PfRh5Nt) and PfRh5FL, but not the C-terminal PfRh5Ct (Galaway, *et al.*, 2017). This study further provided results showing that only PfRh5Nt-coated beads could purify the P113 from parasite culture lysate, in contrast to PfRh5Ct-coated beads. Further AVEXIS assay experiments using a truncated recombinant PfRh5Nt revealed a 19 residue long linear sequence of PhRh5Nt binding a cluster of 14 cysteine residues at the N-terminus of P113 (Galaway, *et al.*, 2017). The findings were consistent with the formation of a tetrameric P113 complex with its C-terminus consisting of a coiled-coil region. However, since the 19-amino acid region of PfRh5Nt is a highly flexible disordered segment, there is always the risk that it behaves as a "sticky" protein due to their increased plasticity and malleability, which would lead to potential false-positive protein-protein interaction results in the experiment performed (Ozbabacan, *et al.*, 2011). Furthermore, none of the recombinant P113 proteins were assessed for correct folding. This should be considered along with the fact that a 2016 paper reports that via a proteomics approach on blood-stage *P. falciparum* parasites, finding P113 to be associated with PTEX (Elsworth, *et al.*, 2016) (aside from being found in the parasitophorous vacuole, PTEX can also be found inside the intracellular dense granules of the merozoite prior to invasion of host erythrocytes (Bullen, *et al.*, 2012)). Finally, other groups have been unable to see what Galaway *et al.* reports when they have carried out experiments (personal communication with Drs. Paul Gilson and Hayley Bullen at the Burnet Institute, April 2017).

This would appear to be revealing a great deal of conflicting results around P113. However, there are several possibilities; firstly, low resolution immunofluorescent assays (IFA) shown in (Galaway, *et al.*, 2017) may be making P113 appear to be on the surface of the merozoite, when in fact it is still in the dense granules of the parasite. Alternatively, P113 may possess dual functions; serving the PTEX complex in addition to providing an anchor for the PAIN-complex via PfRh5 binding. It would not be the first time that a *Plasmodium* parasite protein held a dual function (for example, the RhopH protein complex (Ito, *et al.*, 2017)). A definitive answer should come from further analysis of P113, including further cellular-based assays and ultra-resolution microscopy. However, due to the paucity of confirmation currently and skepticism over published results, this thesis will not regard P113 as part of the PAIN-complex.

CyRPA has also been suggested to be tethered to the merozoite membrane via a GPI-anchor, thus providing the component that holds the PAIN-complex to the parasitic cell surface (as the model in figure 1.6 demonstrates (Reddy, *et al.*, 2015)). However, several contradicting papers have subsequently been published indicating that CyRPA in fact lacks a GPI anchor and is instead a secreted protein (Dreyer, *et al.*, 2012) (Chen, *et al.*, 2017) (Favuzza, *et al.*, 2017) (Volz, *et al.*, 2016). Attempts at replicating the 2015 studies from Reddy *et al.* were carried out by (Volz, *et al.*, 2016) but were unsuccessful, and an experiment with immunoprecipitated CyRPA (one native and one haemagglutinin-tagged) that was analysed via LC-MS/MS revealed an absence of GPI.

The crystal structure of CyRPA, both alone and with inhibitory antibodies bound was revealed by two groups recently; one achieving a resolution of 2.5 Å (Favuzza, *et al.*, 2017), the other to 2.44 Å (Chen, *et al.*, 2017). The crystal structure (figure 1.10) reveals a canonical six-bladed β-propeller fold containing four intra-sheet disulfide bonds (Cys19-Cys35, Cys90-Cys92, Cys151-Cys169 and Cys229-Cys239) and one inter-sheet disulfide bond between Cys274 and Cys298. Each blade of the β-

propeller consists of a four-stranded anti-parallel β-sheet, each of which are connected by loops that have been suggested to have some flexibility in solution. The structure of CyRPA shows similarities to viral and bacterial sialidases (although some key residues for this activity were absent), though no significant sialidase activity was detected when tested (Chen, *et al.*, 2017). CyRPA is also important for being the bridge between PfRh5 and PfRipr, as these two proteins are not directly associated with one another in the PAIN-complex (personal communication with Prof. Alan Cowman, April 2017, and Galaway, *et al.*, 2017).

Figure 1.10. Crystal Structure of *P. falciparum* CyRPA: The ribbon-model of CyRPA that was solved from X-ray diffraction data at a resolution of 2.5 Å. Each blade of the β-propeller is labelled 1-6 starting with the N-terminus in black. Blade 1 consists of a single N-terminal β-strand, plus three C-terminal β-strands (shown in red). Disulfide bonds are shown as yellow ball and sticks. One conformer is shown in white, the other in black loop regions (refer to the arrows). Image and information from Favuzza, *et al.*, 2017.

PfRipr is a cysteine-rich protein (87 cysteine residues in total) expressed as a 126 kDa protein in mature schizonts and stored in the micronemes prior to host cell egress and invasion (Chen, *et al.*, 2011). Here it is processed into two ~65 kDa polypeptides by a yet unknown enzyme, and when these are run on SDS-PAGE gels under both reducing and non-reducing conditions, they appear at the same place, suggestive that intramolecular disulfide bonding is not present between fragments. The opposite is the case when each individual ~65 kDa fragment is run under the same conditions; non-reducing samples have enhanced gel mobility which suggest the intramolecular disulfide bonds contribute to tight folding. PfRipr additionally has no transmembrane region and is peripherally-associated with the cell membrane. The processing of PfRipr most likely occurs in schizonts or as merozoites egress from an infected erythrocyte (before a fresh invasion event), as evidenced by the presence of processed PfRipr in parasite blood stage development. Full-length PfRipr contains ten EGF-like domains, two localized to the N-terminus and eight clustered in tandem near the C-terminus. These EGF-like domains are predicted to contain β-strands as the only secondary structure (as determined by prediction software such as the 1D protein structure prediction server), and are all linked by three amino acids, aside from EGFs 5 and 6, which have an unusual 17 amino acid linker containing two cysteines. Currently however, the three-dimensional structure of PfRipr remains elusive.

Due to the paucity of published data on PfRipr from a protein biochemical perspective (despite being a key protein in the PAIN-complex), it can be assumed that many have tried and failed to produce either a full-length or a partial fragment. There is one publication reporting the use of a plant cell expression system to produce one of the EGF-like domains of PfRipr, EGF6 (Spiegel, *et al.*, 2015). The problem with this work however, was a lack of visualization of the purified protein under non-reducing conditions, in addition to a lack of any biophysical characterization such as circular dichroism (CD) or nuclear magnetic resonance spectroscopy (NMR) to analyse secondary structure and folding. That said, one key strength of the work was in their demonstration of the production of an in-tandem

approach to vaccine development, with various recombinant malaria antigens from several life cycle stages utilized, an approach others in the vaccine development community should take heed of.

As previously mentioned, PfRipr binds to CyRPA, however there is no information in the current literature regarding potential binding regions. The EGF-like domains of PfRipr can be assumed to be potentially involved in mediating protein-protein interactions within the complex as indeed EGF domains in other organisms including other apicomplexan parasites such as *Toxoplasma gondii* are involved in protein complex formation (Meissner, *et al.*, 2002) (Wouters, *et al.*, 2005). However, this remains mere speculation. Despite the relative paucity of biochemical information on PfRipr, its importance as a prospective vaccine candidate was recently further confirmed during the functional profiling of the *Plasmodium* genome (Bushell, *et al.*, 2017). The study utilized the common rodent model of malaria and *P. berghei* knockout mutants and measured competitive growth rates after drug treatment. This revealed the essentiality of the *P. berghei ripr* gene.

1.4.4. Proteins of the PAIN-complex and Calcium Signalling

Following the initial three events of invasion; attachment, erythrocyte deformation and merozoite reorientation, there is a calcium flux into the erythrocyte causing it to shrink, an event that appears to both play a key role in invasion and facilitated (at least in part) by the proteins of the PAIN-complex upon BSG binding (Volz, *et al.*, 2016) (Weiss, *et al.*, 2015). The interaction between erythrocytic BSG and parasitic PfRh5 of the PAIN-complex is associated with Ca^{2+} release and therefore has been suggested as essential in the formation of a pore allowing Ca^{2+} flux between parasite and erythrocyte (Volz, *et al.*, 2016) (Aniweh, *et al.*, 2017). Recently, successful conditional gene deletion research was carried out, reducing the expression of the proteins by 60-95% by using dimerizable Cre recombinase

(DiCre) to generate RiprloxCre and CyRPAloxCre, and growing parasites in rapamycin (Volz, *et al.*, 2016). *P. falciparum* parasites with conditionally deleted *pfripr* and *cyrpa* only expressing PfRh5 were shown to attach to and deform erythrocytes, but invasion was ablated. This study suggests that neither PfRipr nor CyRPA is essential in these *initial* steps, but supports the idea that they are essential for the subsequent internalization of the parasite, potentially by pore formation. This was further supported by the observation that merozoites with conditional *pfripr* and *cyrpa* knockouts did not show the Ca^{2+}spike seen in wildtype parasites at any point during interaction with host cells.

1.4.5. Invasion Inhibition by Antibodies Against the PAIN-complex

Research has been carried out on vaccines targeting the blood-stage of *P. falciparum,* due to the potentially more pragmatic benefits contrasted to those targeting either the sexual stages, or sporozoites. Prospective vaccines targeting these aforementioned stages may interrupt either the transmission cycle or sporozoites, but problematically would not have a direct effect on an established infection in the case of a single sporozoite or ookinete breaking through and continuing the pathogenic replicative cycle (Felgner, *et al.*, 2013) (Crompton, *et al.*, 2010). Attempts to inhibit blood-stage erythrocyte invasion using antibodies to several of the *P. falciparum* complex proteins have so far been relatively successful *in vitro* (reviewed in Crompton, *et al.*, 2010). Additionally, since it is the blood-stage of the infection that actually causes disease, it can be argued that it should be included in vaccine strategies.

A key extensive study had identified PfRh5 and PfRipr as two of the top blood stage targets for vaccine development efforts based on their quality and reactivity to human antibodies, in addition to their association with protective immunity (Richards, *et al.*, 2013). With respect to CyRPA, even prior to its association with the PAIN-complex, it was known as a prospective blood-stage vaccine candidate

following screening of conserved *P. falciparum* proteins that elicited inhibitory antibodies (Favuzza, *et al.*, 2012).

Inhibiting the PfRh5-BSG interaction by rabbit anti-PfRh5 polyclonal IgG showed that merozoite binding and deformation of erythrocytes occurred as normal, but invasion frequency was reduced by around 90% (Weiss, *et al.*, 2015). This was similar to the levels of invasion inhibition demonstrated previously (Douglas, *et al.*, 2014). Another study using serum from rabbits and rats immunized with $rPfRH5_{63}$ found that anti-PfRh5 antibodies blocked erythrocyte binding in a dose-dependent manner, with potent binding inhibition of both the native and recombinant PfRH5 protein (Reddy, *et al.*, 2014). More recently, anti-PfRh5Nt antibodies developed by Galaway, *et al.*, 2017 inhibited invasion and parasite growth *in vitro* in contrast to Douglas, *et al.*, 2014, whose anti-PfRh5 antibodies were directed against an epitope in the N-terminus of PfRh5 (though this may be due to missing key functional epitopes). The most recent work using human-derived anti-PfRh5 antibodies came from the previously mentioned vaccine studies by Payne, *et al.*, 2017. The serum antibodies from the human volunteers that were vaccinated with PfRh5-encoded viral vectors were able to disrupt both the binding between PfRh5 and BSG, but also the formation of the PAIN-complex via the interruption of the interaction between PfRh5 and CyRPA. It is also important to note that antibody titres to PfRh5 have been shown to have a strong correlation with protection against severe clinical episodes of malaria amongst a Papua New Guinean population (Chiu, *et al.*, 2014).

Antibodies against other proteins of the PAIN-complex have also shown good promise as vaccine candidates. It was shown that CyRPA-specific monoclonal antibodies (mAbs) inhibit invasion *in vitro* and in mice engrafted with human erythrocytes (Dreyer, *et al.*, 2012). Further trials showed highly potent invasion inhibition using anti-CyRPAFL antibodies, with an EC_{50} of ~3 mg/mL total IgG and up to 85% inhibition at 10 mg/mL total IgG (Reddy, *et al.*, 2015). 70% invasion inhibition with an anti-CyRPA mAb that reacted to a conformational

epitope was demonstrated in the structural paper by (Chen, *et al.*, 2017). Furthermore, an additive effect was demonstrated when administering both anti-CyRPA and anti-PfRh5 antibodies to humanized mice infected with malaria, with lower levels of parasitaemia developing contrasted to controls (Favuzza, *et al.*, 2017).

Far fewer studies have been carried out on anti-PfRipr antibodies. In part this is due to PfRipr being the youngest of the proteins known in the PAIN-complex (Chen, *et al.*, 2011). The first significant study was carried out in 2011, where the FCR3 strain of *P. falciparum* displayed 80% reduced growth in assays utilizing anti-PfRipr antibodies (Chen, *et al.*, 2011). This study also demonstrated that when a combination of anti-PfRipr and anti-PfRh5 antibodies were used, a synergistic effect against parasite growth and invasion occurred amongst various strains of *P. falciparum*. These results were further solidified when additional research demonstrated anti-PfRipr antibodies provided potent erythrocyte invasion inhibition against 3D7 (~63% invasion inhibition) and FVO (~77% invasion inhibition) strains of parasites in a GIA (Ntege, *et al.*, 2016). Finally, there is a strong association between individuals with high IgG3 antibody subclass levels against both PfRipr and PfRh5, and a reduced risk of malaria, suggestive of concluding that IgG3 responses against these invasion ligands are important for mediating naturally-acquired immunity against clinical malaria (Weaver, *et al.*, 2016). However, this study was focused on individuals in Papua New Guinea, and so it will be important to carry out similar studies on other malarious regions around the globe.

1.5 Conclusions

This thesis introduction highlights the immense complexity of the malaria parasite and the challenges we face in trying to eradicate this great scourge. It should go without saying that the global eradication of malaria is highly likely to help usher in

further human prosperity based on the established bi-directional link between poverty and malaria. Recent research has shed more light on the PfRh family of proteins such as PfRh4 and the essential PAIN-complex including PfRh5, CyRPA, and PfRipr; all of which are promising potential vaccine targets. The least is currently known about PfRipr, with a dearth of current structural and binding site information, in addition to immunological data, all things that should be addressed to aid better understanding of the protein complex binding. Progress towards a clearer image of the full complex and its binding sites will be essential for a successful structural vaccinology approach to create a vaccine using these essential proteins. It should be clear that going forwards, an important vaccine strategy must involve targeting combinations of antigens expressed at the multiple life-cycle stages, in addition to targeting the PAIN-complex proteins.

Ideally, it needs to be established which parts of these proteins are both essential to PAIN-complex formation, invasion and can give rise to inhibitory antibodies, so that a protein engineering approach is taken to develop an in-tandem or in combination mix to be tested in vaccination trials (similar to the work carried out by Speigel, *et al.*, 2015). Further, it would behove the malaria vaccine community to also test this alongside and in combination with RTS,s-like vaccines and others corresponding to other life-cycle stages.

1.6 Thesis Aims

The aims of this doctoral thesis are to contribute to an important body of knowledge around several PfRh-family proteins including PfRh4, PfRh5 and the difficult to study PfRipr. During the course of this research it emerged that PfRipr was a *P. falciparum* protein of great importance and so the bulk of efforts in this thesis is dedicated to an in-depth biophysical chemical analysis of this protein. Furthermore, methods for the industrially-applicable scaling up of anti-malaria vaccine components was of importance due to the simple fact that it will do the

world little good to produce a vaccine whose distribution would be inhibited by an inability to scale-up its production in a financially appropriate manner. With that in mind, the specific aims of this thesis are:

1. Development of a set of mAbs directed against PfRh4, and attempts at elucidating both the structure of PfRh4 and its CR1 binding region.

2. The bacterial production of recombinant PfRipr protein fragments, their biophysical characterization, attempts at structural analysis via crystallography, and binding studies against PfRh5.

3. The expression of PfRipr fragments in *Pichia*, their scaled-up production, biophysical characterization, and attempts at both elucidation of the binding region of PfRipr against CyRPA, in addition to NMR and crystallography-based structural analysis.

CHAPTER 2

General Methods.

"I have nothing to offer but blood, toil, tears and sweat"

Winston Churchill, 1940.

2.1 *E. coli*-Based Cloning

In silico cloning using DNA Strider software were used to ensure that restriction enzyme digests would be compatible for when the synthetic genes were to be ligated into their final expression vectors, and to ensure all genes remained in-frame. Where required, software on www.resitefinder.appspot.com was used to search for silent restriction enzyme sites. Synthetic genes were manufactured by Integrated DNA Technologies (IDT) or Geneart where specified. The plasmid maps for the common expression vectors used in this thesis are shown in figures 2.1 and 2.2.

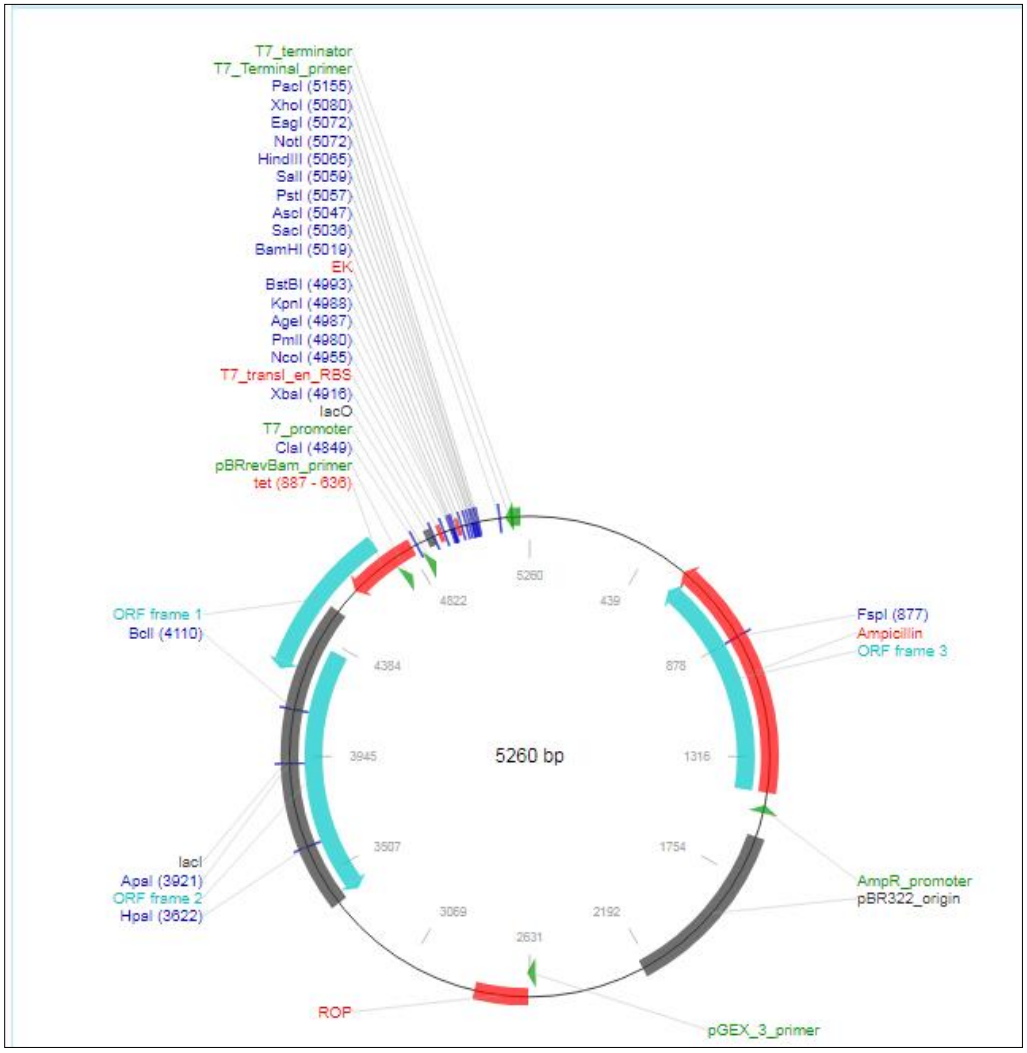


Figure 2.1. Plasmid Map for pET45b(+): Synthetic *P. falciparum* genes of interest were inserted into multiple cloning site, typically via *Bam*HI and *Xho*I.

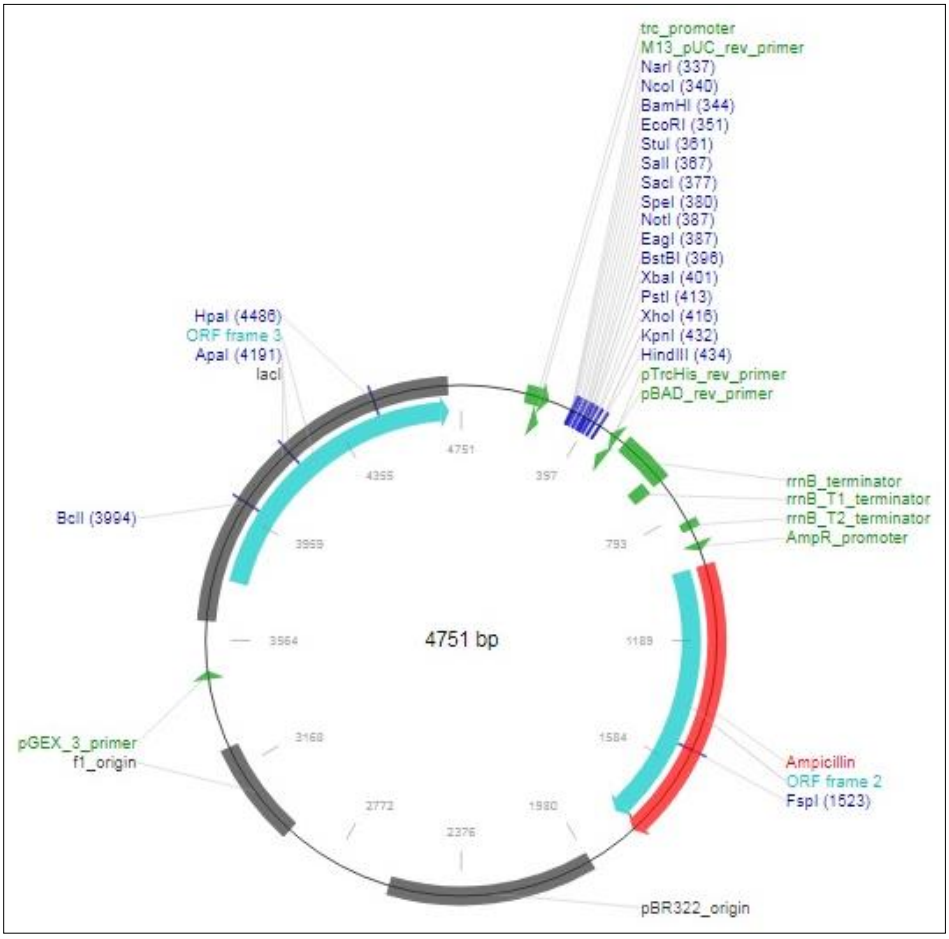


Figure 2.2. Plasmid Map for pProExHTA: Synthetic *P. falciparum* genes of interest were inserted into multiple cloning site, typically via *Bam*HI and *Xho*I.

Lyophilized synthetic genes were reconstituted in 20 – 50 µl of TE buffer, then a 1:100 dilution prepared using double-distilled H_2O (ddH_2O) to give a concentration of 1 ng/µl. 4 – 5 µl of the diluted DNA was added to 20 µl of electrocompetent DH10β *E. coli* cells and left to incubate on ice for 30-minutes. Tube contents were transferred to a pre-chilled 0.1 cm electroporation cuvette and electroporation carried out using the following settings on a Bio-Rad Gene Pulser II: 1250 V, 25 Ohms, 200 µF. Cuvette contents were transferred to sterile tubes containing 0.6 ml of S broth then incubated at 37 ºC for one-hour at 200 rpm. 100 µl of the culture was plated onto 50 ml LB agar plates containing 50 µl of 100

mg/ml carbenicillin (or 40 mg/ml kanamycin where required). A negative control was also prepared using H_2O only. All plates were grown overnight at 37 ºC.

Resulting colonies were picked and inoculated into 3 ml of S broth containing 3 µl of 100 mg/ml carbenicillin (or 40 mg/ml kanamycin where required) and left overnight in a 37 ºC shaking incubator (200 rpm). A miniprep-based DNA extraction was carried out according to the protocol given in the PureLink™ Quick Plasmid Miniprep Kit (ThermoFisher). DNA concentration was calculated via NanoDrop.

Restriction enzyme digests were prepared by mixing up to 30 µl of plasmid with 3 – 5 µl of appropriate NEB buffer (New England Biolabs), 5 µl of 10 x bovine serum albumin (BSA), 7 µl of ddH_2O and 1.5 µl of each restriction enzyme. The reaction mixtures were incubated for 2.5 hours at 37 ºC, then 5 µl of 10 x dephosphorylation buffer and 1 µl of alkaline phosphatase was added to the tubes as required. Samples were analysed via agarose gel electrophoresis using an 0.8 – 1.2% agarose gel prepared in 0.5 x TBE (or TAE were required) containing 0.01% ethidium bromide (following addition of 4 – 6 µl of 6 x loading dye to each sample) and run for 80-minutes at 65 V. 5 µl of Generuler 1 kb molecular weight marker (ThermoFisher) were used for size comparisons. Gels were soaked in 0.5 x TBE (or TAE were required) with 0.02% ethidium bromide for one-hour and analysed via UV visualization, with bands of interest extracted via scalpel blade. Gel slices had DNA extracted and purified vie the methods outlined in the Invitrogen Purelink® Quick Gel Extraction Kit.

A ligation reaction between the expression vector backbone and the synthetic gene insert was carried out by mixing 2 µl and 6.6 µl respectively of the aforementioned DNA, 1 µl of 10 x ligation buffer, and 0.4 µl of T4 ligase (a negative control was also set up substituting ddH_2O instead of the synthetic gene insert), then incubating for 17 hours at 16 ºC to generate the final expression plasmid.

Next, the DNA was concentrated via a phenol chloroform extraction. This involved adding 40 µl ddH2O to the ligation reaction tube, followed by 50 µl of phenol chloroform and tube agitation via vortex mixer. Sample tubes centrifuged at 10 000 × g for four-minutes, supernatant removed and added to fresh tubes containing 50 µl of phenol chloroform, followed by a repeat of this step. The final supernatant was added to a tube containing 1 µl of glycogen, 5 µl 3 M sodium acetate (pH 5.2), and 130 µl 100% ethanol. The tube was centrifuged at 10 000 × g for 30-minutes, supernatant removed and 140 µl of 80% ethanol added to the resulting pellet. Tube was vortexed, centrifuged at 10 000 × g for 30-minutes, supernatant removed and the pellet air dried before dissolving in 5 µl of ddH_2O.

Where ethanol precipitation was used, the following protocol was carried out- one tenth the sample volume of 3 M sodium acetate (pH 5.2) and 2.5 x sample volume of 100% ethanol added to DNA sample tube, which was incubated at -70 ºC for ten-minutes. After incubation, sample centrifuged at 10 000 × g for 15-minutes at 4 ºC. Removed supernatant and added 2.5 x sample volumes of chilled 80% ethanol then gently vortexed. The sample tube was centrifuged at 10 000 × g for 15-minutes at 4 ºC, supernatant removed, then re-centrifuged for a further five-minutes. After air-drying, the resulting pellet was resuspended in 5 – 10 µl of ddH_2O.

Transformation and plating of DH10β *E. coli* carried out using the final expression plasmids with subsequent miniprep and NanoDrop as described above. Using 1 µl of the confirmed expression plasmids, a 1:100 aliquot was prepared in 99 µl ddH_2O, with a 1 µl and a 10 µl volume used for the transformation of 40 µl tubes of electrocompetent BL21 gold *E. coli* cells (Agilent technologies) via the aforementioned methodology.

2.2 Production of Electrocompetent BL21 Gold *E. coli* Cells

Production of electrocompetent BL21 gold *E. coli* cells involved making a streak plate from a glycerol stock then picking a colony into 6 ml of S broth to be left shaking at 200 rpm overnight at 37 ºC. The culture was added to 600 ml of S broth and grown until the OD_{600} was 0.4, then the culture flask transferred to wet ice in a cold room for 30-minutes, with gently agitation halfway through the incubation. 99

Flask contents centrifuged at 4 ºC for ten-minutes at 3500 × g. Cells resuspended in 600 ml chilled ddH_2O and left to incubate on wet ice in the cold room for one-hour. Flask contents re-centrifuged as before and washed twice more in chilled ddH2O. Pellet was resuspended in an equal volume of 10% glycerol, chilled on wet ice for 30-minutes, centrifuged, and then re-washed in 10% glycerol. After careful decanting of supernatant, the pellet was resuspended in 6 ml of 10% glycerol. 40 µl aliquots were prepared and stored at -80 ºC until needed.

2.3 Glycerol Stock Preparation

Glycerol stocks were prepared by taking 10 ml of an overnight culture grown in 100 ml of L broth + 100 µl of 100 mg/ml carbenicillin (37 ºC shaking at 200 rpm), and centrifuging at ~1200 × g for ten-minutes. After the supernatant was discarded, the pellet was resuspended in 10 ml of L broth and re-centrifuged. The supernatant was discarded, being replaced with a 15% glycerol mix (sterile and prepared in L broth), with the pellet gently resuspended and snap frozen in liquid nitrogen for storage at -80 ºC.

2.4 SDS-PAGE Protein Analysis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was carried out by adding an appropriate volume of sample (generally to be able to visualize 2 μg of protein) with up to 30 μl of 2 x NuPAGE™ LDS sample buffer (SB) and 6 μl of dithiothreitol (DTT, thus making reducing SB, or RSB) then boiled for up to five-minutes at 95 ºC. For NI and I samples, taking into account the dilution, the amount of broth required in μl to set aside as the I sample was calculated by using the formula $150/OD_{600}$ (this may also be used for the NI sample if the OD_{600} reading is significantly over 0.6). These samples were analysed on a 4-12% bis-tris gel (Invitrogen) via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) run at 150 V in a 5% MOPS buffer for 80-minutes (and subsequent gel staining in Coomassie blue buffer prepared as described in (Lawrence & Besir, 2009)). Unless otherwise mentioned, all molecular weight markers used are Precision Plus Protein Unstained Standards 10 – 250 kDa range (BioRad).

2.5 Small-Scale Test Expression

Note, this section is adapted from Haggarty-Weir *et al.*, 2017.

After plating freshly transformed *E. coli* containing the final expression plasmid onto LB agar with a suitable selective marker (i.e. kanamycin), an individual colony was inoculated into a 50 - 100 ml flask (baffled where possible) with 10 ml of S broth plus antibiotic, growing overnight at 37 °C with agitation at 220 rpm. 300 ml baffled flasks each with 30 ml of S broth plus antibiotic were prepared and 1 ml of the starter culture was inoculated into each and then placed in a 37 °C incubator at 220 rpm. These cultures were grown until an optical density at 600 nm (OD_{600}) of 0.6 – 0.8, then were split the flasks so one-third remained at 37 °C, one-third at 30 °C one-third in an 18 °C incubator (for the latter two incubators, the flasks

from the 37 °C incubator were removed when the OD_{600} was just below 0.6 so as to pre-equilibrate the temperature before induction). Of each three lots of flasks, one of each were induced with 0.25 mM, 0.5 mM and 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) after taking a 1 ml non-induced (NI) sample. The flasks at 37 °C were grown for four hours, with a 1 ml post-induction (I) sample taken, harvesting the rest by centrifugation (8 – 10 000 × g for 10-minutes). The flasks at 30 °C and 18 °C were grown for 8 and 16 hours respectively before collecting the I sample, measuring the final OD_{600} and harvesting the flask contents.

250 µl and 50 µl of each NI sample and I sample (respectively) was taken, centrifuged at 10 000 × g for one-minute, supernatant removed then 10 µl of 1 x RSB added and SDS-PAGE carried out. Each sample showing evidence of induction for recombinant protein production was centrifuged as described and resuspended in 2 ml of lysis buffer (50 mM Tris pH 8, 250 mM NaCl, 0.1% triton X-100, 2 mM EDTA, spatula tip of lysozyme and 1.5 µl of DNaseI), then processed via sonication (90 sec on/100 sec off at 35% amplitude for 7.5-minutes) or mechanical cell disruption. 50 µl of the sample post-lysis was centrifuged as described above to achieve separation of the supernatant from the pellet. 6 µl of the supernatant had 10 µl of 1 x RSB added, and for the pellet, 50 µl of water was added, resuspension carried out thoroughly, then 6 µl was taken and added to 10 µl of 1 x RSB. SDS-PAGE analysis on these samples was carried out to determine if any protein was expressed in the inclusion body fraction. Note that scaled-up production of proteins typically used 500 ml of S or L broth in 3000 ml baffled flasks.

2.6 Processing of Scaled-Up Bacterial Cultures

Flask contents were harvested via centrifugation at around 10 000 × g for ten-minutes at 4 ºC. After the supernatant was decanted, beakers were placed on ice

and the pellet was resuspended in 160 ml of lysis buffer (400 mM NaCl, 40 mM Tris pH 8, 100 mM PMSF, 2 mg/ml DNaseI per 40 ml, and 1 Roche cOmplete ULTRA EDTA-free protease inhibitor cocktail tablet per 40 ml). Using a cell disruptor (Constant Systems Ltd) with pressure set to 26 kPa, cells were lysed by two rounds of cell disruption. A 50 µl post-lysis sample was taken and centrifuged at 10 000 × g for one-minute. Around 7 µl of the supernatant was taken and added to 26 µl of 1 x RSB (thus making the post-induction supernatant sample), with the rest of the supernatant decanted. 50 µl of water was added to the resultant pellet which was resuspended and 7 µl removed to add to 19 µl of 1 x RSB to make the post-induction pellet fraction. These samples were prepared for SDS-PAGE analysis as previously described. The lysed bacterial stock was then centrifuged at 10 000 × g for 30-minutes at 4 ºC.

2.7 Nickel-Affinity Purification of His_6-Tagged Proteins

The supernatant of the lysed material was then applied to 4 ml of a 50% nickel-NTA (Ni-NTA) agarose bead slurry (from Qiagen), mixed and allowed to settle, and then the flow-through collected. 75 ml of Ni-wash buffer (5 mM imidazole, 400 mM NaCl, 40 mM Tris pH 8) was applied and collected, then 10 ml of Ni-elution buffer (500 mM imidazole, 400 mM NaCl, 40 mM Tris pH 8) was applied with 2 ml fractions collected. The concentration of the protein in the eluted fractions was determined with an A280 nm NanoDrop 1000 Spectrophotometer (NanoDrop) using elution buffer as the blank. 2 µl samples from the flow through, Ni-wash and Ni-elution fractions were added to 10 µl of 1 x RSB and analysed via SDS-PAGE as previously described.

40 µl of a Roche cOmplete™ protease inhibitor solution (1 tablet in 1 ml of sterile water + 0.5 mM EDTA) was added to each eluted fraction. Based on SDS-PAGE gel results, key fractions were combined and added to a pre-washed and equilibrated (with sterile water and Ni-elution buffer, respectively) 30 kDa cut-off

protein concentration tube (Merck Amicon Ultra-15 Centrifugal Filter Units). Sample was centrifuged at 3500 × g for ten-minute lots at 4 ºC until ~2 ml was left. The NanoDrop was used to calculate final protein concentration and to determine the amount of sample to run on an SDS-PAGE gel to visualize ~2 µg of protein.

2.8 Western Blot Methodology

Western transfers were carried out at 100 V for 60-minutes on ice onto a nitrocellulose membrane, which was subsequently blocked with 10 % blocking solution (10% skim milk in PBS with 1% Triton-X, PBS-T) for one-hour at room temperature. Membranes were washed thrice with PBS-T. A 1:200 dilution of anti-His_5 antibody (Qiagen) was prepared in 1% blocking solution and added to the membrane. These samples were left gently rocking at room temperature for one-hour, then were washed twice with five-minute washes in 1% blocking solution. A secondary antibody solution of 1:1000 anti-mouse HRP antibodies were prepared in 1% blocking solution and added to the membranes for one-hour, gently rocking at room temperature. This was followed by five five-minute wash steps and gentle dry-blotting of the membranes. 10 ml of enhanced chemiluminescence ECL solution (Amersham) was applied with rigorous mixing for three-minutes followed by a one-second exposure and development on X-ray film.

2.9. General Chromatography

Concentrated protein samples were centrifuged at 10 000 × g for ten-minutes at 4 °C to clear particulates before injecting the protein onto a size-exclusion chromatography (SEC) column (GE Superdex HiLoad 200 16/600 pg) run on an ÄKTA purifier system (GE). Flow rate was set at 1 ml/min with 0.8 ml fractions

collected and the buffer used was 20 mM MES + 75 mM NaCl, pH 6 (also used to equilibrate the column prior to sample loading). Key fractions had the protein concentration determined via NanoDrop and 2 μg visualized via SDS-PAGE.

The key fractions were combined and applied to a 5 ml Mono S ion exchange chromatography column (GE 5/50 GL) run at a flow rate of 0.5 ml/min (collecting 0.5 ml fractions) and using buffer A (20 mM MES pH 5.5) plus an increasing gradient of buffer B (20 mM MES pH 5.5 + 1 M NaCl). After SDS-PAGE analysis of 2 μg of each fraction (concentration determined via NanoDrop), another pre-washed and equilibrated protein concentration tube was used for concentrating key fractions down to 0.5 – 1 ml.

CHAPTER 3

Investigations into PfRh4.

"To live is to suffer, to survive is to find some meaning in the suffering"

Friedrich Nietzsche, *Thus spake Zarathustra*, 1891.

3.1 Overview

This study entailed the production and purification of recombinant truncated proteins corresponding to ~89 kDa PfRh4 proteins (PfRh4.9; includes ten alanine scan mutants from amino acids 282 - 342 called $PfRh4_{6Ala1-10}$), an 82 kDa deletion mutant lacking amino acids 283 – 341 (termed $PfRh4.9_{\Delta 58}$), and a 43 kDa truncation from amino acids 104 – 465 called PfRh4.18. These proteins were selected to isolate the binding site on PfRh4 that interacts with its cognate erythrocytic receptor, CR1. PfRh4.18 was selected due to being a small truncation possessing the alleged binding region to CCP1-3 of CR1 (Tham, *et al.*, 2009), and so it made sense to try and produce it for crystal trials. The rationale was to give atomic level detail needed to have definitive elucidation of the binding interface.

To hedge against crystal trial failure and any difficulties with production of PfRh4.18, the series of alanine scan mutants and the deletion mutant $PfRh4.9_{\Delta 58}$ (as previously mentioned covering the supposed binding region in PfRh4.19 (Tham, *et al.*, 2009)) was to be developed for binding studies against CCP1-3. These would be used in ELISA assays to look for diminished levels of binding overall compared to native PfRh4.9. If a particular mutant/s from $PfRh4_{6Ala1-10}$ showed a significantly diminished level of binding to CCP1-3 contrasted to PfRh4.9, then an estimate of the amino acid residues involved could be ascertained. This would open up the door to developing further mutants to further narrow the binding site down. However, if no difference in binding affinity was to be detected then this would be suggestive that the predicted binding region based on the inhibitory antibody screening in Tham, *et al.*, 2009 was inaccurate.

The production of anti-PfRh4.9 mAbs was important given that they could be used to screen for the blocking of binding to CR1, and for their use in the identification of inhibitory epitopes in the binding interface. Those mAbs which were not inhibitory to the interaction between PfRh4.9 and CR1 could be used as tools to aid co-crystallization trials. Furthermore, all mAbs which showed high

affinity binding to PfRh4.9 could be screened for tight binding to PfRh4.18 with a view to use them for crystal trial screening of this protein alone.

3.2 Production and Purification of PfRh4.9 for Vaccination Studies

3.2.1 Methods for the Production and Purification of PfRh4.9

Glycerol stocks of PfRh4.9-containing BL21 (DE3) *E. coli* were produced as described in Tham, *et al*., 2009, with proteins produced and purified as detailed in Lim, *et al*., 2015.

SDS-PAGE performed as described in section 2.4, cultures were processed as described in section 2.6, and Ni-affinity purification carried out as per section 2.7. Following Ni-affinity purification, SEC and ion-exchange chromatography were carried out according to section 2.9 to generate a final, pure sample of PfRh4.9.

3.2.2 Methods for PfRh4.9 Vaccination Studies

Methods were carried out as described (Lim, *et al*., 2015). It should be noted that the author also carried out parasite culturing, Western blots, immunofluorescent assays, and competition ELISAs (not mentioned in the Lim, *et al.,* 2015 paper in the Appendix). The ELISA method used by the author is outlined below.

A competition ELISA was developed to screen the anti-PfRh4.9 mAbs using 96-well flat-bottomed plates (Maxisorp; Nunc). Wells were coated with 100 µl of 2 µg/ml CCP1-3 (provided by Dr. Christoph Q Schmidt of Ulm University), or in the case of the controls, PBS only, and 2 µg/ml PfRh4.9. After overnight incubation at 4 °C, plates were washed 3 times with PBS with 1% Tween 20 (PBS-

T) and blocked with 200 µl of 5% skim milk blocking buffer for one-hour at room temperature. During this step, in a regular 96-well plate, 100 µl of 2 µg/ml PfRh4.9 was mixed with 100 µl of mAb supernatant supplied by the WEHI antibody facility (for the control media and the positive control polyclonal antibody supernatant, 100 µl of a 1:50 dilution was used), and left for one-hour at room temperature. Following another 3-part PBS-T wash, the ELISA plate had the non-ELISA plate contents containing the PfRh4.9+antibody mixes (and controls) added for a two-hour incubation at room temperature. Note, after one-hour, 100 µl of 2 µg/ml PfRh4.9 was added to the CCP1-3 control well, and blocking buffer added to other control wells. Further, 50 µl of mouse anti-PfRh4 mAb 10C9 (1:1000 made up in 5% blocking buffer) was added to the PfRh4.9 control and CCP1-3+PfRh4.9 wells for one-hour at room temperature. After 3 × PBS-T washes, 50 µl of anti-mouse HRP antibodies (1:2000 made in 5% blocking buffer) were added to wells for one-hour at room temperature, followed by 3 more PBS-T washes. 100 µl of an HRP-substrate reagent 3,3,5,5-tetramethylbenzidine (TMB, ThermoFisher) was added to each well until the background signal appeared, at which time 50 µl of 2 M phosphoric acid were used to quench the reaction. Absorbance was measured at 450 nm.

3.2.3 Results for the Production and Purification of PfRh4.9 and Vaccination Studies

See the paper in the Appendix under Lim, *et al.,* 2015. Additional results not included in the paper are thus: An ELISA was performed to ensure mAb reactivity to recombinant PfRh4.9, whereby ELISA plates were coated with PfRh4.9 then exposed to a different mAb, followed by detection (Appendix table A1). A competition ELISA was done to determine which anti-PfRh4.9 mAbs were potentially capable of ablating the interaction between PfRh4.9 and CCP1-3 (figure 3.1). This identified six potential mAbs; 3A6, 9G9, 5H12, 4C5, 2A7 and 2H9, with the latter two demonstrating very prominent levels of CCP1-3 binding inhibition.

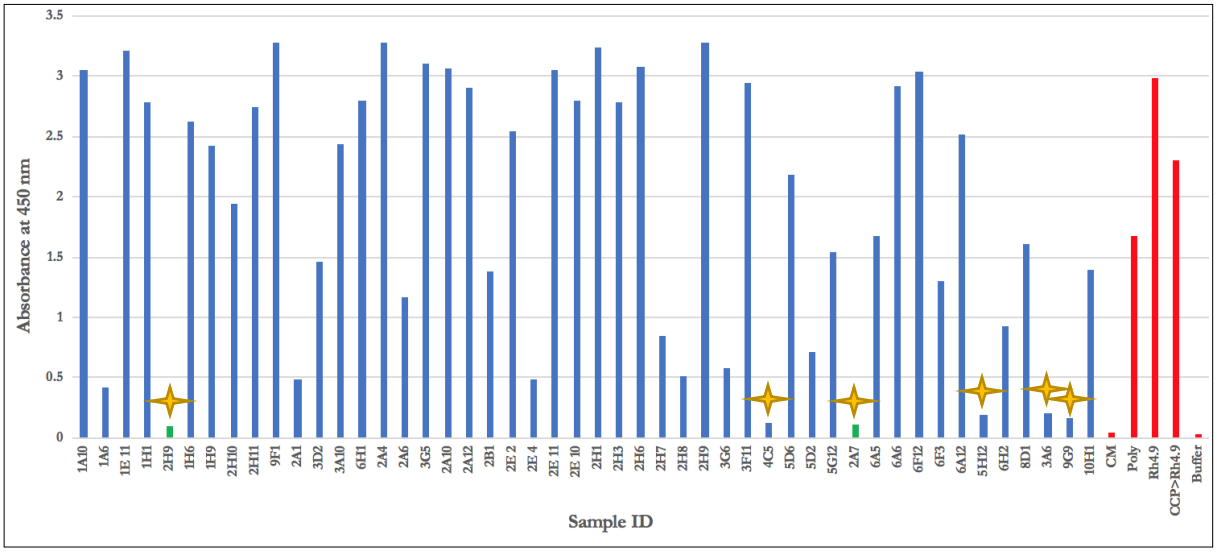


Figure 3.1. Competition ELISA Screening of αPfRh4.9 mAbs: Detection of PfRh4.9 with mouse anti-PfRh4 mAb 10C9 by ELISA following co-incubation of PfRh4.9 with various mAbs and their application to CCP1-3-coated wells. Bound antigen-antibody complexes detected with anti-mouse HRP. Green and blue bars are the mAbs screened (green are samples taken for further experimentation), red bars are controls, yellow stars are the key inhibitory mAbs. Abbreviations: CM- negative control media, Poly- polyclonal antibody supernatant AB185 #1232 from fusions (positive control), Rh4.9- PfRh4.9-only coated well (positive control), CCP>Rh4.9- positive control for CCP1-3 + PfRh4.9 interaction, Buffer- buffer-coated well negative control.

3.3 Cloning and Production of PfRh4.18 for Structural Characterization

3.3.1 Methods for the Cloning and Production of PfRh4.18

Cloning, protein test expression, SDS-PAGE, Ni-affinity purification and Western blot analysis were all carried out as per section 2. Specific details and changes are given below. The pET45b plasmid was used for expression by cloning the synthetic gene for PfRh4.18. DNA Strider software was used to ensure a *Bam*HI and *Xho*I restriction enzyme digest was compatible with the synthetic gene going from pET45b into pProExHTA (containing an N-terminal His_6-polyhistidine tag and tobacco etch virus, TEV, enzyme cleavage site) whilst the PfRh4.18 gene

remained in frame. The pProExHTA (called ProEx throughout the rest of this thesis) was donated by Mr. Tony Triglia. The ligated plasmid of pProExHTA and PfRh4.18 will be referred to as ProEx+PfRh4.18.

DNA sequencing was carried out by Applied Genetic Diagnostics, The University of Melbourne. 1 µl of each primer used at a concentration of 5 µM, and a 5 µl sample of DNA (100 ng/kb) to be sequenced was provided. The primers were:

WH99 (reverse)- GCGctcgagttattacttgatattggtgcattgcagcatcaggtcg

WH100 (forward)- CGCggatcccctgaaaaccgatatcatcctgaact

3.3.2 Results of the Cloning and Production of PfRh4.18

Following plasmid sequencing, transformed BL21 gold cells were used for 4 x 2 L cultures split between IPTG concentrations ranging from 0.05 mM to 1 mM. Flasks were grown at 37 °C and induced at 30 °C. The NI, I, pellet and supernatant samples were kept for analysis. SDS-PAGE results show a strong band around 43 kDa in the pellet fractions of samples induced with 0.25, 0.5 and 1 mM IPTG. The Western blot showed a weak band in the I and pellet samples of the 0.05 mM IPTG-induced sample, and a positive band in the I, supernatant and pellet samples of the 1 mM IPTG-induced sample.

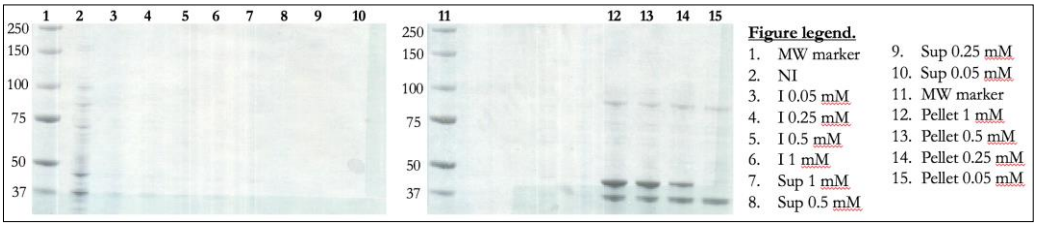


Figure 3.2. SDS-PAGE of Large-Scale Test Expression of PfRh4.18: 15 µl of each sample was loaded onto their respective wells on the gel. The post lysis pellet and supernatant samples here were obtained after using a cell disrupter (rather than sonication).

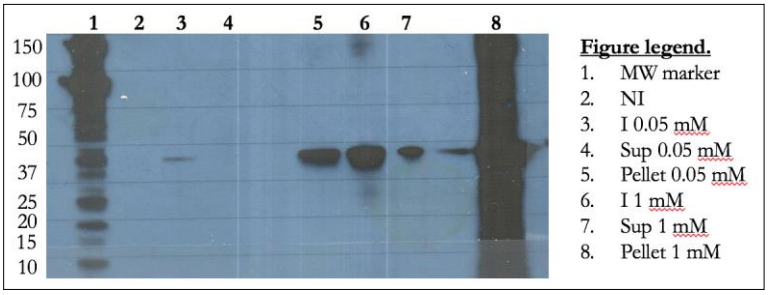


Figure 3.3. Western Blot of Large-Scale Test Expression of PfRh4.18: A 1 in 10 dilution of the samples that were prepared for the SDS-PAGE were used for the Western blot. The primary antibody was Qiagen anti-His_5 mouse antibody, the secondary antibody used was anti-mouse HRP. Development time of the Western blot was five-seconds.

Next, the effect of dropping the induction temperature (from 30 to 18 °C) on PfRh4.18 expression was carried out. An IPTG concentration of 0.3 mM was used, and samples from both temperatures were taken and analysed via Western blot (figure 3.4). Whilst there was a negligible effect on protein band intensity for the I samples of each induction temperature, the lower temperature did appear to shift a higher proportion of PfRh4.18 into the insoluble pellet fraction.

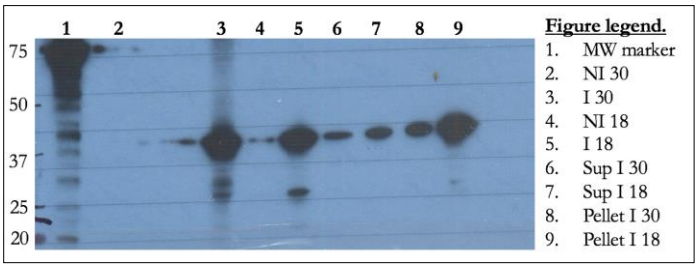


Figure 3.4. Western Blot of PfRh4.18 Comparing IPTG Induction at 18 °C and 30 °C: 1.5 µl of each sample was run. Temperature of induction (either 18 or 30 °C) is given next to the sample names in the lane legend. The primary antibody was anti-His_5 mouse antibody, the secondary was anti-mouse HRP. Development time of the Western blot was five-seconds.

Attempts at refolding from inclusion bodies in the pellet of PfRh4.18-expressing cells that were reconstituted from glycerol stocks were unsuccessful (methodology for the conditions given in Haggarty-Weir, *et al.*, 2017). This may have been due to low quality glycerol stocks, so it may have been better to use the ProEx+PfRh4.18

DNA to carry out a fresh transformation. However, production of PfRh4.18 for structural studies was discontinued due to supervisory interference.

3.4 Cloning, Production and Purification of PfRh4.9 Mutants for CR1 Binding Studies

3.4.1 Methods for the Cloning, Production and Purification of PfRh4.9 Mutants for CR1 Binding Studies

Hydropathy plots generated with Kyte and Doolittle hydropathy parameter software using the key 58 amino acid sequence at the N-terminus of PfRh4 identified in Tham, *et al.*, 2009, flanked by 21 amino acids either side as shown below (with the key 58 amino acid region in red):

KVNEYINNSDCHLTCSKYKTIINNYIDEIITTNTNIYENKYNLPQERIIKNYNHNGINND
DNFIEYNILNADPDLRSHFITLLVSRKQLIYIEYIYFINK

Restriction enzyme digests of the pET45b+PfRh4.9 plasmid for assessing silent restriction sites (Tham, *et al.*, 2009) were carried out as previously described in section 2.1, but using 1 – 5 µl of pET45b+PfRh4.9, 1 µl of appropriate endonuclease buffer (10 x NEB2, 10 x NEB4 or *Eco*RI buffer), 0.3 – 1 µl of restriction enzyme (*Hind*III, *Bam*HI, *Xho*I, *Mfe*I and/or *Eco*RI), 1 – 1.5 µl 10 x BSA, and ddH_2O as appropriate.

Agarose gel electrophoreses carried out as previously described in section 2.1, with the additional use of 5 µl of 100 bp DNA ladder (ThermoFisher). DNA Strider software was used to assess *in silico* cloning. Note that pET45b was used to allow for a hexa-histidine tag to be placed on the N-terminus of each protein.

DNA Strider was used to design the PfRh4.9 synthetic gene containing the *Hind*III and *Mfe*I restriction sites, in addition to designing each mutant for the alanine scan mutagenesis. 11 constructs were also developed including a PfRh4 region from residues 154 – 365 but without the key 58 amino acid region containing residues 283 – 341 (PfRh4.9$_{\Delta 58}$; flanked by *Hind*III and *Mfe*I for cloning into the new PfRh4.9 construct), and 10 genes corresponding to residues 282 – 342 with each lot of six residues mutated to alanines to produce PfRh4.9$_{6Ala1\text{-}10}$. These additional synthetic genes also contained the *Hind*III and *Mfe*I restriction sites.

DNA sequencing of samples was carried out as described in section 3.3.1, using the below primers for the new pET45b+PfRh4.9 mutants (WHT 79 and 82 used for the deletion mutant, WHT 79 and 80 used for the 6Ala$_{1\text{-}8}$ constructs, and WHT 81 for 6Ala$_{9\text{-}10}$):

WHT79 (forward) CGCggatccCagcaaagaaaaga

WHT80 (reverse) GCGactcgagttattaaaaatgagaacgcagatccg

WHT81 (forwards) CGCggatccCatcgacagtgaaaacgagaagc

WHT82 (reverse) GCGctcgagttattaaatctcgttcagcttattcagga

Protein expression and Western blotting was carried out as per sections 2.5 and 2.8 respectively. SEC methodology remained the same as outlined in section 2.9, except 1 ml fractions were collected. Protein concentrators were used to buffer exchange from the 20 mM MES pH 6.5 + 100 mM NaCl used in SEC to 1 × PBS with 75 mM NaCl (pH 7.5) prior to ELISA setup.

A concentration ELISA using 96-well flat-bottomed plates (Maxisorp; Nunc) was carried out to determine the amount of mutant PfRh4.9 protein to use for the CCP1-3 binding assay. Plates coated with 150 µl of 16 µg/ml of each PfRh4.9 mutant protein (and regular PfRh4.9) in triplicate, with a 1:3 serial dilution carried over into the next five wells. Three wells had 100 µl of 1 × PBS added as a negative control, then all plates incubated overnight at 4 °C. Wells washed, then blocked for one-hour at room temperature with 200 µl of 5% skim milk made in

PBS-T. Washing followed by 50 µl of 1:2000 10C9 mouse anti-PfRh4 mAb 10C9 (made in 5% blocking buffer) added to each well for one-hour. After washing, 50 µl of 1:2000 anti-mouse HRP antibodies (made in 5% blocking buffer) added for one-hour followed by 3 x PBT-T washes and 2 × PBS washes. 100 µl of an HRP-substrate reagent 3,3,5,5-tetramethylbenzidine (TMB, ThermoFisher) was added to each well until the background signal appeared, then 50 µl of 2 M phosphoric acid quenched the reaction. Absorbance was measured at 450 nm.

The protein-protein interaction ELISA was set up by incubating triplicates of 100 µl of 2 µg/ml CCP1-3, triplicates of 4 µg/ml (100 µl) of each PfRh4.9 protein, and triplicates of 100 µl PBS-only into 96-well flat-bottomed plates (Maxisorp; Nunc) overnight at 4 °C. After 3 × PBS-T washes, 200 µl of 5% blocking solution was added to each well and incubated at room temperature for two-hours. Plates washed, then a 1:2 dilution series of PfRh4.9 and mutant proteins (previously prepared in regular 96-well plates in a volume of 100 µl starting at a concentration of 16 µg/ml, and made in triplicate) were added to their respective CCP1-3 and buffer-only control wells (all other control wells had 100 µl of PBS added to them). After a two-hour incubation at room temperature and washing, 50 µl of 1:1000 10C9 mouse anti-PfRh4 antibodies (made in 5% blocking buffer) was added to each well for 1.5-hours at room temperature. After washing the plates, 50 µl of 1:2000 anti-mouse HRP antibodies (made in 5% blocking buffer) added to each well for one-hour at room temperature, then washed with PBS-T and 2 x washes with PBS. Detection was carried out as previously described above.

3.4.2 Results: Cloning, Production and Purification of PfRh4.9 Mutants for CR1 Binding Studies

Following on from reports of a potential 58 amino acid region within PfRh4 that is involved in binding CR1 (Tham, *et al.*, 2009), a hydropathy plot of this region was

generated and analysed to assess hydrophobicity of the region and thus if it is more or less likely to be surface exposed (figure 3.5). Residues 283 – 341 were highly hydrophilic and therefore possessed a lower likelihood of being a region buried in a hydrophobic core, thus being indicative of a region of the protein that is surface exposed. The spike around position 65 – 70 corresponds to Arg329.

Next, potential unique silent restriction sites had to be identified up- and downstream of residues 283 – 341 that would be compatible with engineering and cloning into a pET45b vector without changing the reading frame for PfRh4.9 expression. This would become important for the cloning in of the alanine scan mutant genes of PfRh4.18 designed.

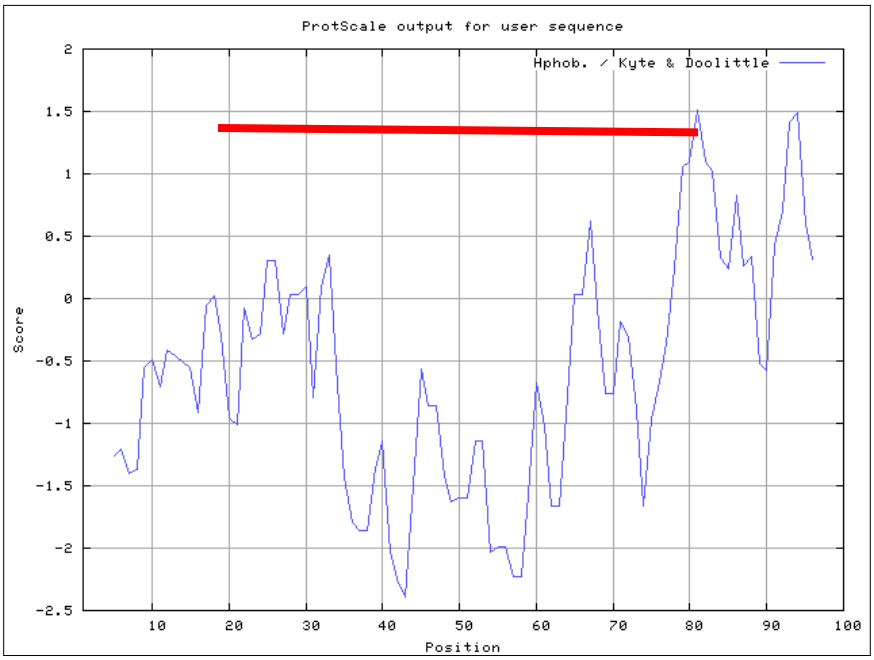


Figure 3.5. Hydropathy Plot of PfRh4 Residues 262 – 362: The red line indicates the key 58 amino acid region of PfRh4 (residues 283 – 341) that is potentially involved in CR1 binding. A negative score correlates to increasing hydrophilicity and a positive score correlates to increasing hydrophobicity.

Four prospective sites were identified with *Hind*III and *Mfe*I at the 5` and 3` regions respectively looking the best in terms of distance from the key 58 amino

acid region of the gene. To confirm that these sites would not lead to enzymatic digestion of the genes outside of the designated restriction sites, a series of restriction enzyme digests were carried out on the pET45b+PfRh4.9 plasmid and analysed via agarose gel electrophoresis (figure 3.6). The pET45b+PfRh4.9 plasmid size is 7526 bp (containing the 1819 bp PfRh4.9 gene), and so if there was activity of both the *Hind*III (or MfeI; neither of these sites should be present in the plasmid) and the secondary enzyme (*Bam*HI or *Xho*I; both sites present in the plasmid), then 2 bands would be expected. However, a single band was seen in all sample wells, indicating that the prospective silent restriction sites would be suitable to use.

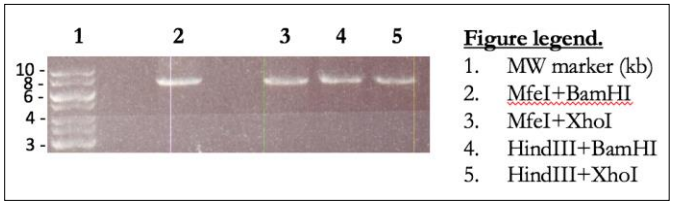


Figure 3.6. Agarose Gel Electrophoresis of Restriction Enzyme Digested pET45b+PfRh4.9: Single bands around 7.5 kb appear in all lanes. Gel is 0.8% agarose in 0.5 x TBE buffer.

Following transformation of DH10β cells with the new IDT plasmids (pIDT), a colony count was performed revealing no colonies on the negative control plate, 64 colonies for the new PfRh4.9, 118 colonies for the deletion mutant, and between 200 and 1000 for the 10 alanine mutants. After a DNA miniprep, each new synthetic gene was excised from the pIDT for ligation into pET45b. After preparation of the ligated new pET45b+PfRh4.9 vector containing the *Hind*III and *Mfe*I restriction sites, the plasmid was used to transform DH10β cells, yielding colonies contained the pET45b+PfRh4.9 plasmid. To confirm the presence of the new restriction sites, the DNA from the colony yielding the highest concentration (86 ng/µl) was analysed via gel electrophoresis following a *Mfe*I/*Xho*I and *Hind*III/*Xho*I restriction digest (figure 3.7). The DNA sequencing also revealed confirmation of the pET45b+PfRh4.9 vector (results not shown).

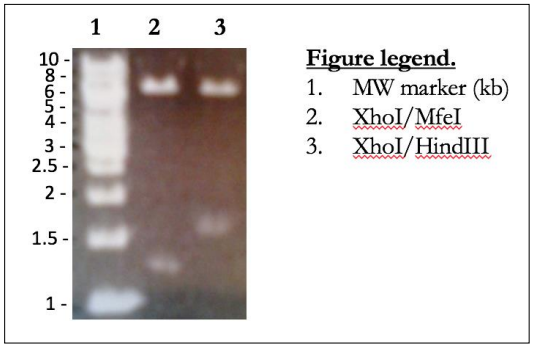


Figure 3.7. *Xho*I and *Mfe*I or *Hind*III Restriction Digest of pET45+PfRh4.9: The *Xho*I/MfeI double digest shown in lane 2 reveals two bands, one around 6.1 kb, the other around 1.3 kb, as expected. The *Xho*I/*Hind*III double digest in lane 3 shows a band around 5.8 kB and a lower band around 1.6 kb, as expected. Samples were run on a 0.8% agarose 0.5 x TBE gel.

Next the new pET45b+PfRh4.9 plasmid underwent a double digest using *Mfe*I and *Hind*III to prepare for the ligation of $6Ala_X$ mutants, and the deletion gene. DNA was gel purified, concentration determined (43 – 75 ng/µl in a 10 µl volume), then analysed via agarose gel electrophoresis (figure 3.8). The remaining DNA was used for ligating into the *Mfe*I/*Hind*III-cut pET45b+PfRh4.9 plasmid, then electroporated into DH10β cells and plated to yield 250 colonies. After minipreps of selected colonies, the DNA was used for a *Xho*I and *Not*I-Hf restriction digest and sequencing (results not shown).

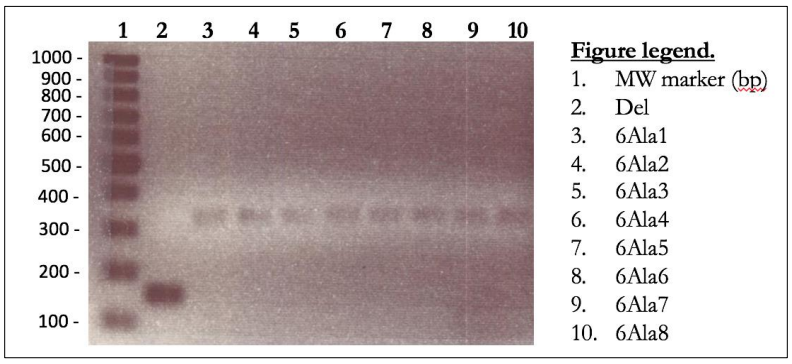


Figure 3.8. Concentration Gel of 6Ala and Deletion Mutant Genes: 2 µl of each sample was analysed. 6Ala1 – 8 mutant genes are the expected ~330 bp, and the deletion gene appeared at the expected ~160 bp (6Ala9 and 6Ala10 were also run on a separate gel but not shown here). 1.2% agarose gel prepared in 0.5 x TAE buffer, run at 70 V for 75-minutes.

Following the production and transformation of electrocompetent BL21 gold cells with each plasmid, a colony count was carried out revealing thousands of colonies for all transformants, and zero in the negative control (water only). Test expression following cell growth at 37 °C and IPTG-mediated induction at 30 °C was carried out to analyse the NI, I, pellet and supernatant samples for each expressed protein. This was examined by Western blot (figure 3.9), revealing all proteins expressed at around the same levels, with all pellet samples appearing at a slightly higher intensity than the supernatant fractions. Next, for assessing the impact of induction temperature on protein expression, the previous text expression experiment was carried out but now inducing at 18 °C. SDS-PAGE analysis of the samples was carried out (figure 3.10). Inducing at 18 °C appeared to shift a very slight amount of the protein expression into the soluble supernatant fraction, leading to a roughly equal level of protein found between here and the pellet.

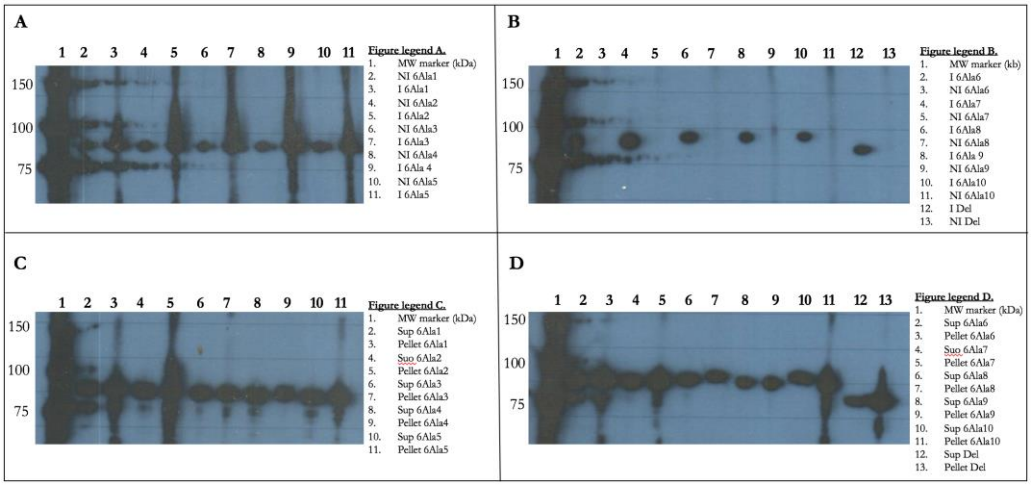


Figure 3.9. Western Blot of PfRh4.9$_{6Ala1-10}$ and PfRh4.9$_{\Delta 58}$ Expression: A. Proteins appeared at the expected ~90 kDa size, even in the NI samples (likely due to "leaky" expression), with more expression visualized in I samples. B. As for A, except the NI samples were over-boiled. C. Pellet and Sup samples for PfRh4.9$_{6Ala1-5}$ show increased levels of protein as inclusion bodies in the pellet fraction. D. As for C, but shows PfRh4.9$_{6Ala6-10}$ and PfRh4.9$_{\Delta 58}$, with the latter protein appearing at the expected size of ~82 kDa. Primary antibodies used were anti-hexaHis mouse antibodies, with anti-mouse HRP used for the secondary. Development time of the X-ray film was three-seconds.

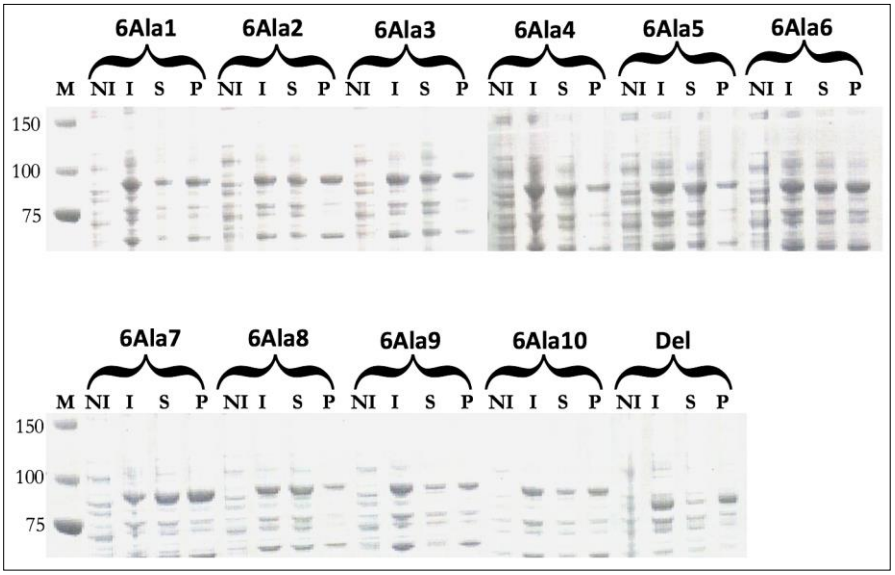


Figure 3.10. SDS-PAGE Results of PfRh4.9$_{6Ala1-10}$ and PfRh4.9$_{\Delta 58}$ Test Expression: All samples demonstrate the appearance of a significant band around 90 kDa (or ~82 for the deletion mutant) following induction with 1 mM IPTG (at 18 ºC), as expected. Abbreviations used: M- molecular weight marker (kDa), N- non-induced, I- induced, P-pellet, S-supernatant.

Scale-up expression of each mutant protein in 600 ml of culture medium was carried out. Whilst the samples were roughly analogous to what was seen in terms of protein expression and division between pellet and supernatant, the PfRh4.9$_{\Delta 58}$ was highly predominant in the pellet fraction (data not shown). This may be the result of using half as much IPTG for the induction, as this was the only variable.

The supernatant post-cell lysis was Ni^{2+}-affinity purified, and 5 x 2 ml elutions collected (revealing approximately 1 mg of protein in each fraction). Fractions were combined and concentrated to under 5 ml for application to SEC. 2 µl of each concentrated sample was viewed on a gel via SDS-PAGE revealing the presence of some contaminating proteins, many of which could be removed during SEC (data not shown). Key fractions were concentrated (figure 3.11); however, PfRh4.9$_{\Delta 58}$ was unable to be detected via SDS-PAGE or Western blot (data not shown). Further attempts at producing PfRh4.9$_{\Delta 58}$ were unsuccessful and due to time consideration, it was dropped from further analysis.

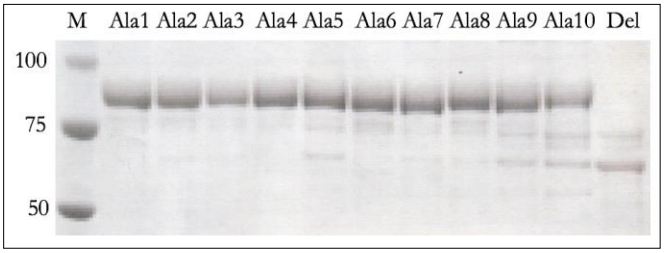


Figure 3.11. SDS-PAGE of Concentrated PfRh4.9 Mutants: Around 3 μg of the final concentrated proteins were run under reducing conditions revealing the presence of mostly monomer, with small levels of degradation products present.

After performing a concentration ELISA (results not shown) with each protein to determine that anti-PfRh4 antibodies could detect the proteins and to evaluate the starting concentration for each, a protein-protein interaction ELISA was set up to screen the interaction of each mutant protein against CCP1-3, and contrasted to PfRh4.9 binding to CCP1-3 (figure 3.12). All mutant interactions were very similar to native PfRh4.9 binding to CCP1-3, thus suggesting the binding region of PfRh4 to be outside of the amino acid region covering residues 283 – 341.

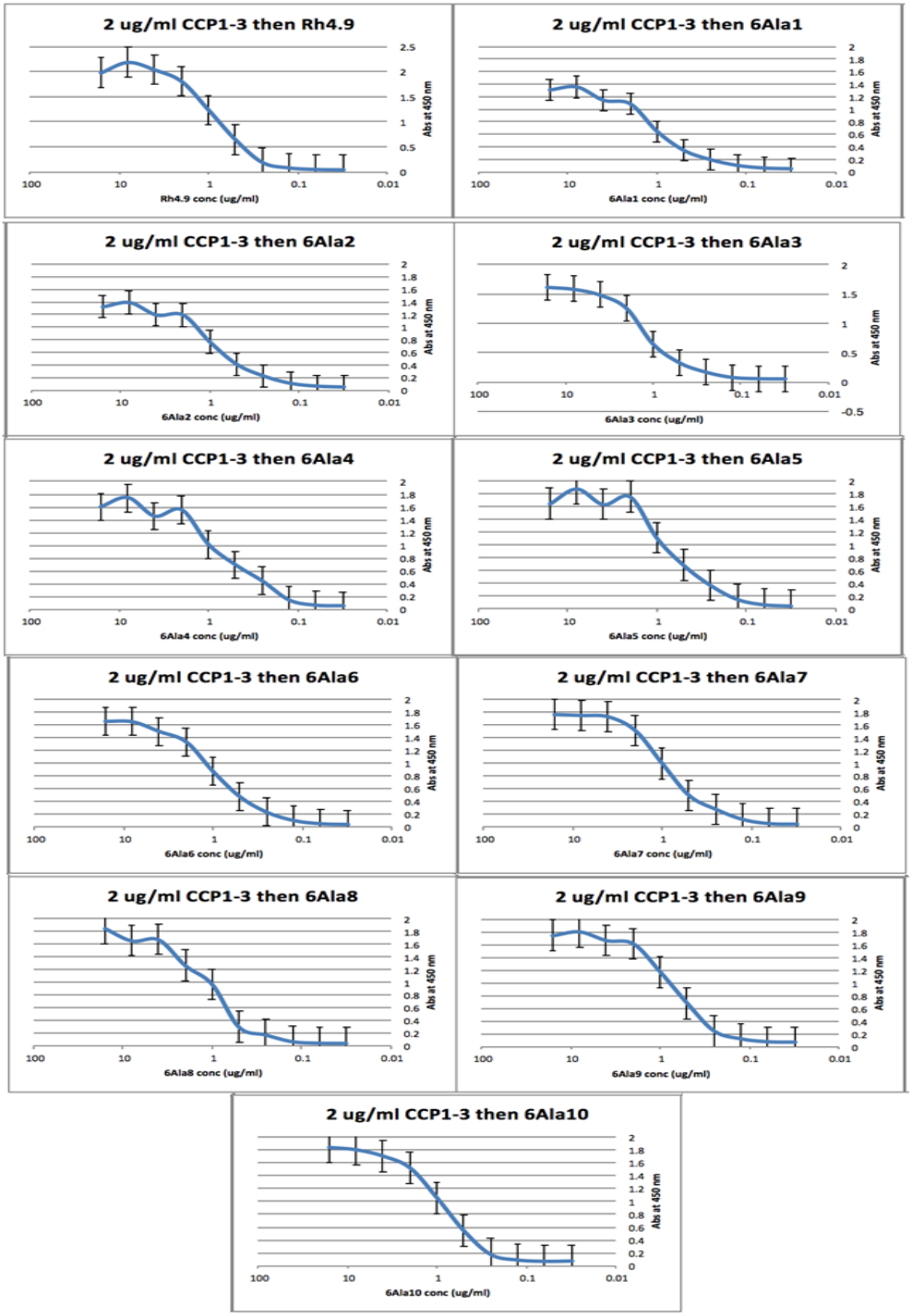


Figure 3.12. PfRh4.9 Alanine Scan Mutagenesis: PfRh4.9 mutant proteins were screened for binding against CCP1-3. Data shown are from triplicate experiments. The buffer control gave an absorbance (at 450 nm) of 0.05, CCP1-3-only control was 0.05, protein-only background was 0.15, and positive controls for each protein gave a reading of 1.7. Error bars are standard deviation.

3.5 Discussion and Conclusions

Until recent evidence came to light demonstrating the essentiality of the PAIN-complex proteins (Volz, *et al.*, 2016) (Bushell, *et al.*, 2017) and their stability (Chen, *et al.*, 2014) (Chen, *et al.*, 2017), PfRh4 was a candidate for vaccine research due to its function in sialic acid-independent invasion of erythrocytes via CR1 (Tham, *et al.*, 2010), and the correlation between protective antibodies against it in a PNG population and protection against severe malaria (Richards, *et al.*, 2013). Additionally, for an understanding of fundamental parasite biology, the discovery of inhibitory mAbs to PfRh4 is useful. For this reason, the study carried out had aimed to produce functionally important mAbs against PfRh4 using a recombinantly-produced truncation, PfRh4.9, in addition to developing a smaller, more stable truncation (known as PfRh4.18), and using alanine scan mutagenesis to elucidate the precise region of PfRh4 involved in binding CR1.

The production of PfRh4.9 revealed an unstable protein despite efforts to prevent degradation including the use of protease inhibitors and cold storage. The instability of recombinant PfRh4 has been reported elsewhere in the literature (Richards, *et al.*, 2013), although one may not have been unable to initially tell due to SDS-PAGE results shown closely cropped to the protein band of interest in various publications such as (Gaur, *et al.*, 2007). This of course is a significant issue for vaccine design as component stability is essential (reviewed in Kumru, *et al.*, 2014). However, despite the instability of PfRh4.9, a suitable amount of mostly monomeric antigen was produced and used for Balb/c mouse vaccinations. Both Western blot and IFA analysis (figure 1 of Lim, *et al.,* 2015) demonstrated the sera from vaccinated mice to be reactive against PfRh4, as expected. Although, what was unexpected was the finding that the antibodies found in the competition ELISA in figure 3.1 that prevented PfRh4.9 binding to CCP1-3 (mAbs 2A7 and 2H9), were not found to be inhibitory (and thus not reported on) in the experiments carried out in Lim, *et al.*, 2015 with the mAbs directed against PfRh4.9. 5H12, the most significant inhibitory mAb discovered in Lim, *et al.*, 2015

was almost selected as a potential mAb by the competition ELISA in figure 3.1, but due to time considerations was not selected for further development. It is possible that the discrepancy between the two results arose from human error due to inexperience, though follow-up experiments replicating those carried out in the published paper but using mAbs 2A7 and 2H9, are warranted.

The development of the 5H12 mAb represents the first anti-PfRh4 neutralizing antibody reported in the literature (Lim, *et al.*, 2015). Utilizing three PfRh4-CR1 interaction assays based on ELISA, FRET, and immunoprecipitation, it was shown that 5H12 is capable of disrupting the PfRh4-CR1 interaction. Addition of 5H12 mAbs ablates PfRh4 binding to erythrocytes and inhibits the PfRh4-CR1 pathway in follow-up *P. falciparum* GIA experiments reported in figure 3.C – E of Lim, *et al.*, 2015. 5H12 would also be an ideal mAb to use in crystal trials that make use of recombinant PfRh4 (such as PfRh4.18), as this would potentially allow for accurate conformational epitope mapping of the residues responsible for binding to the paratope present on CR1 (akin to the strategy used to map the epitope region of PfRh5 (Chen, *et al.*, 2014), and CyRPA (Chen, *et al.*, 2017)).

PfRh4.18 was a smaller protein than PfRh4.9, and it was selected for development in order to pursue crystallography-based structural studies. Cloning of the gene resulted in the transformation of protein-expressing BL21 gold *E. coli* cells. A variety of expression conditions were screened, ultimately revealing two potential expression pathways utilizing bacterial cells grown at 37 °C then induced with 0.3 mM IPTG at 18 °C. Both of which appeared to produce initially very pure protein (figure 3.4, lanes 6 – 9). The next step would be to scale-up to several litres culture (to overcome low protein expression) and to obtain soluble protein after harvesting the *E. coli* (as shown in the Western blot in figure 3.4, lane 7). Alternatively, the use of refolding from inclusion bodies could be utilized given the presence of high levels of protein in the cellular pellet fraction (figure 3.4, lane 9).

Initial test expressions had begun utilizing the latter expression pathway but failed to yield protein. This may have been due to having a poor-quality glycerol stock, and so it was planned to re-transform BL21 gold cells with ProEx+PfRh4.18 plasmid. Before this was executed in full, the author's involvement in the PfRh4 project was discontinued. In the future, one would be recommended to carry out refolding trials of PfRh4.18 utilizing various refolding strategies. Alternatively, one could consider using a different expression system, such as insect cells, which have proven useful for PfRh5 and CyRPA (Chen, *et al.*, 2014) (Chen, *et al.*, 2017).

To probe the potential binding site on PfRh4 that is responsible for interaction with CR1, an alanine scan mutagenesis strategy was implemented. This involved changing sets of six amino acids at a time to alanine residues from the section of the PfRh4 protein covered within residues 282 – 342. It was decided to include the flanking amino acids from the 58-amino acid region (283 – 341) identified in Tham, *et al.*, 2009 in order to have an easy division of ten mutant constructs. Whilst the plan was to also include PfRh4.9$_{\Delta 58}$, a suitable quantity of this was unable to be produced in time for the protein-protein interaction ELISA. The results of said ELISA (figure 3.12) do not reveal any significant aberration of binding between CCP1-3 and the titration of each PfRh4.9 protein. Some of the minor decreases seen for example in the PfRh4.9$_{6Ala1}$ panel are explained by the minor time differences in HRP-TMB reaction quenching. If any of the mutants did give rise to an absence of binding, the results would be of more significant levels. Given the fact that a recombinantly-produced 58 amino acid peptide comprising residues 283 – 341 of PfRh4 fails to bind CR1 (personal communication, Dr. Lin Chen, 2013), combined with the results reported here, it is likely that the binding region lies elsewhere. It is a possibility that when the 58-amino acid region was found via antibody screening, the antibodies were sterically hindering the actual binding site to CR1, and not binding directly to the PfRh4 amino acids that directly interact with CR1. In the future, the ideal method of elucidating this binding region, would be to produce recombinant PfRh4 and co-crystalize with CCP1.

CHAPTER 4

Investigations into Bacterially-Produced PfRipr

"Optimism is the madness of insisting that all is well when we are miserable"

Voltaire

4.1 Overview

Once further evidence came out regarding the importance of PfRipr and PfRh5 as lead vaccine targets (Richards, *et al.*, 2013) (Chiu, *et al.*, 2014) (Wright, *et al.*, 2014) (Volz, *et al.*, 2016) the overall aims of the research shifted away from PfRh4 to focus on the first proteins found in the PAIN-complex. This was of particular importance given the dearth of structural and biochemical information published on PfRipr. This research intended to produce bacterially-expressed PfRipr recombinant protein (with a predominant focus on the EGF-like domains), in addition to production of PfRh5 from an established insect cell expression system (Chen, *et al.*, 2014), in order to assess binding between these two proteins, and to screen PfRh5 against potential inhibitory small molecules.

E. coli was selected to produce the PfRipr truncations as soluble protein due to it remaining a staunch workhorse for producing recombinant proteins ever since its first use by Genentech to develop somatostatin in 1976 (Itakura, *et al.*, 1977). In fact, up to 30% of biopharmaceuticals, 50% of commercial proteins and >70% of proteins produced in research settings are made in *E. coli* (Bill, 2014). However, due to issues that can arise in bacterially-produced recombinant proteins, especially those in possession of disulfide bonds (leading to issues of aggregation for example; reviewed in Rosano & Ceccarelli, 2014), refolding of the EGF-like domains of PfRipr from inclusion body products was to be explored. The advantage of producing recombinant protein from inclusion bodies is that higher yields can be attained and often with far less contaminating bacterial protein (reviewed in Rosano & Ceccarelli, 2014). For this, a refolding methodology was developed followed by extensive biophysical characterization via mass spectrometry, CD, thermal shift and NMR spectroscopy. The latter three methods are of particular importance for secondary structure information for the evaluation of correct folding, which is particularly important when refolding recombinant proteins.

CD spectropolarimeters measure the difference in absorbance between the left and right circularly polarized elements of a protein, often reporting the results in terms of ellipticity of degrees (Kelly, Jess, & Price, 2005). CD signals arise when radiation absorption occurs, and with respect to proteins, the key chromophores are the peptide bonds, aromatic amino acids and their respective sidechains and disulfide bonds, thus allowing spectral bands to be assigned to distinct structural regions of a given protein. Given the unique physical-chemical environment of secondary structure elements, both the presence of α-helix and/or β-sheet, and their approximate quantity can be divulged from analysis of CD spectra. Therefore, CD can be utilized as a tool for providing a level of evidence that one does indeed have correctly folded protein if you know how much of each secondary structure you would approximately expect to see. However, unless using synchrotron radiation to allow for analysis of the sample in the far UV region (below 180 nm, which provides optimal secondary structure information), the secondary structure element estimates will be less reliable and so an additional biophysical method is required.

NMR is a good companion to CD analysis of protein samples, especially for assessment of folding when dealing with a protein that is non-enzymatic in function (whereby the optimal assessment of folding would be substrate-to-product activity assay). NMR measures magnetic properties of atomic nuclei relevant to those found in proteins (including ^{1}H, ^{15}N, and ^{13}C), which possess a spin; meaning they can precess around a magnetic field (Receveur-Bréchot, Bourhis, Uversky, Canard, & Longhi, 2006). When these nuclear spins absorb energy, they are left in an excited state, which over time return to equilibrium; that is, the resonance phenomenon. The nuclear spins create an intrinsic magnetic field, and the resonance frequency reflects the local magnetic field, therefore providing information on the local chemical environment, and hence, the secondary structure of a given protein.

In addition to protein refolding and biophysical characterization, attempts at producing protein crystals of the recombinant PfRipr proteins was a priority given the paucity of structural information in the literature. Finally, screening of anti-PfRh5 and anti-BSG antibodies was carried out in order to identify inhibitory antibodies that could be used as general tools and for crystallography. The strategy for the production of each truncated PfRipr protein is shown in figure 4.0. The progress for each protein made in *E. coli* is shown below in table 1.

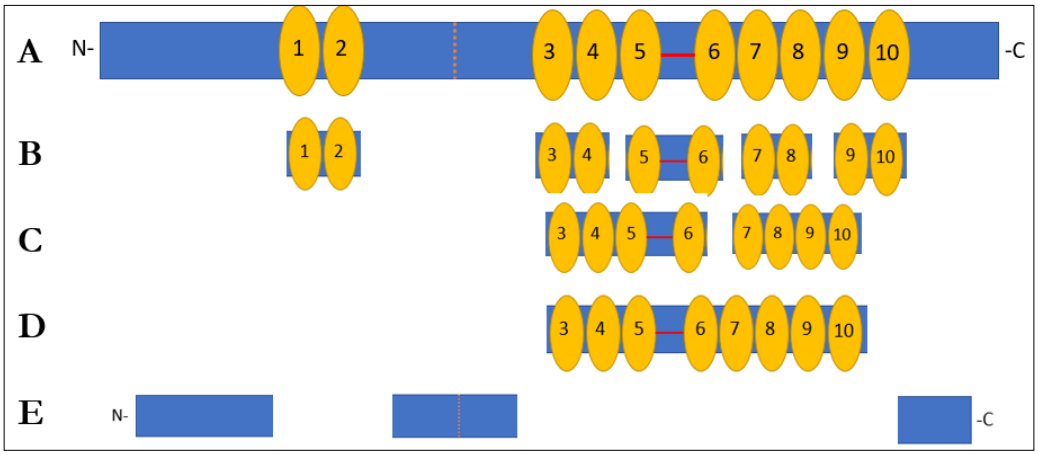


Figure 4.0. Strategy for the Recombinant Production of PfRipr: A. The native PfRipr protein with EGF-like domains numbered in yellow, non-EGF domains shown in blue (with the N and C-termini labelled), the long linker between the 5th and 6th EGF-like domains in red and the polypeptide processing region dotted in orange in the middle of the protein. B. Doublet EGF-like domain truncations. C. Sets of four EGF-like domains in tandem. D. The entire C-terminal region of tandem EGF-like domains. E. Non-EGF domain regions. It should be noted that two forms of the middle non-EGF domain region of PfRipr were to be developed; one with a polyhistidine tag on the N-terminus and one with the tag on the C-terminus. The production of all non-EGF domain truncations of PfRipr was ultimately a failure.

Protein	Purification	MS	WB	TSA	CD	Free Cys Check	1D NMR	2D NMR	Rh5 Binding Assay	CyRPA Binding Assay
EGF1-2	Y	Y	Y	N	N	N	N	N	Y	Y
EGF3-4	Y	N	Y	N	N	N	N	N	Y	N
EGF5-6	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
EGF7-8	Y	N	Y	N	N	N	N	N	Y	N
EGF9-10	Y	N	Y	N	N	N	N	N	Y	N
EGF3-6	Y	Y	Y	Y	Y	N	Y	N	Y	N
EGF7-10	Y	Y	Y	N	N	Y	Y	N	Y	N
EGF3-10	Y	Y	Y	N	N	Y	N	N	Y	N
Ripr21-297	N	N	Y	N	N	N	N	N	N	N
Ripr362-636ct	N	N	Y	N	N	N	N	N	N	N
Ripr978-1086	N	N	Y	N	N	N	N	N	N	N

Table 1. PfRipr Protein Progress Log (*E. coli*): The yellow highlighted regions indicated where a particular examination type was carried out. Abbreviations used: MS, mass spectrometry; WB, Western blot; TSA, thermal shift assay; CD, circular dichroism; Cys, cysteine; NMR, nuclear magnetic resonance spectroscopy.

4.2 Cloning, Production and Purification of PfRipr and PfRh5 Proteins

4.2.1 Methods for Cloning, Production and Purification of PfRipr and PfRh5 Proteins

In general, the cloning methodologies followed that described in section 2.1. *Bam*HI/*Xho*I double digests of pMA-T plasmids were prepared containing each PfRipr EGF construct (EGFs 1-2, 3-4, 5-6, 7-8, 9-10, 3-6 and 7-10) and each non-EGF construct (Ripr21-287, Ripr 362-636, Ripr362-636$_{ct}$, and Ripr978-1086). These were ligated into pET45b+ vectors and also in the case of EGF5-6 only, a pProExHTA vector. Transformation of DH5α and BL21 gold *E. coli* cells,

minipreps, DNA agarose gel electrophoresis and nucleic acid purification were performed as described in section 2.1.

Test expression and purification of each PfRipr construct, was carried out according to the methods outlined in section 2.5. SDS-PAGE analysis and preparation of glycerol stocks were carried out as outlined in section 2.4 and 2.3, respectively. Initial test expression of proteins was carried out in 50 ml bacterial cell cultures with 1 mM IPTG at 30 °C (rather than 18 °C due to incubator availability), and cells were harvested after five-hours. Scaled-up expression of PfRipr EGF-like domains was carried out in a 1.2 L bacterial cultures (unless otherwise noted) incubated at 37 °C until an OD_{600} of 0.6. At this point in the cell growth induction with 1 mM IPTG at 18 °C for 16-hours occurred.

A seed stock of P3 virus containing the PfRh5 synthetic gene in a pgpHFT plasmid that was used to co-transfect Sf21 insect cells with FlashBAC (Oxford Expression Technologies) was donated by Dr. Lin Chen. Protein expression and purification was carried out according to methods in (Chen, *et al.*, 2014).

Western blot analyses were carried out according to the protocol listed in section 2.8. Protein concentration was carried out using 5 or 10-kDa cut-off protein concentrator spin columns (GE Vivaspin ®) for double and quadruple-EGF containing proteins.

Sample filtration was utilized prior to SEC by using a 0.22 μm Millex-GP syringe filter unit (Merck Millipore). SEC was carried out for PfRipr proteins as described in section 2.9, but using either a GE Superdex HiLoad 75 16/600 or 200 16/600, and collecting 1 ml fractions in HTPBS buffer, pH 7.5.

High performance liquid chromatography – elecrospray ionization/mass spectrometry (HPLC-ESI/MS) was carried out by the Bio21 Mass Spectrometry facility in Melbourne.

4.2.2 Results: Cloning, Production and Purification of PfRipr and PfRh5 Proteins

Initial cloning and transformation of Dh5α *E. coli* with pMA-T plasmids containing the PfRipr EGF-like synthetic genes yielded over 1000 colonies on ampicillin enriched agar plates. Following a miniprep and *Bam*HI/*Xho*I double digest, each sample was analysed via gel electrophoresis (data not shown). This produced two sets of bands at the predicted sizes for both the pMA-T vector backbone and the smaller synthetic PfRipr EGF-like genes. The latter were excised, gel purified, and checked on a concentration gel prior to cloning (data not shown). The synthetic genes were each successfully ligated into pET45b+ plasmids (and a pProExHTA plasmid for another construct of EGF5-6) and used to transform DH5α cells.

All DH5α transformation plates yielded 300 – 400 colonies; two to four of which were picked and used to produce cell cultures and subsequent minipreps. The minipreps for each colony yielded 49 – 189 ng/µl of DNA. These were subjected to a *Bam*HI-HF/*Sca*I-HF double digest to confirm expected band sizes on an agarose gel (figure 4.1). In all cases, major bands of the predicted sizes of ~4 kb and 1.6 – 1.8 kb were visible, in addition to some higher bands that were indicative of incomplete enzymatic digestion of the plasmids. BL21 gold cells were transformed and plated, producing >1000 colonies for each electroporated sample carrying double EGF-like genes, and ~200 colonies for cells electroporated with quadruple EGF-like genes.

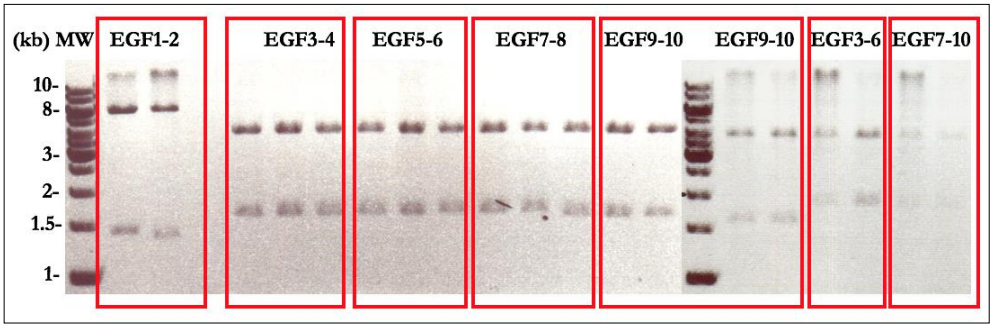


Figure 4.1. Agarose Gel Electrophoresis of pET45b+PfRipr Plasmids: 0.8% agarose (0.5 x TBE) gel revealing two predominant band sizes of ~4 kb and 1.6 – 1.8 kb, in addition to some incompletely digested product. The DNA from each different colony miniprep is found in each box.

Test expression revealed bands of approximate size for I fractions when analysed on SDS-PAGE, with some proteins such as EGFs 1-2, 5-6, 7-8 and 9-10 running slightly higher than expected compared to the molecular weight markers under reducing conditions (figure 4.2 A). This may be due to a variety of factors such as protein molecule isoelectric point (pI) and the amount of SDS bound to the molecule. Induction at 30 °C with 1 mM IPTG revealed different levels of protein expression in pellet and supernatant fractions depending on the protein (figure 4.2 B). A Western blot was carried out using anti-penta-His antibodies (as the proteins all possessed an N-terminal polyhistidine tag) to confirm what was seen in the SDS-PAGE (figure 4.2 C and D). A significant difference appeared to be observed with the I sample of EGF3-4, which appeared to run significantly higher in the Western blot compared to the SDS-PAGE. However, this difference was not observed for the supernatant fraction, likely due to the I sample potentially not being properly sonicated and/or reduced (indicated by the significant smearing appearing towards the top of the I and NI Western blot). All EGFs aside from EGF1-2 and EGF7-8, appeared to predominantly be associated with the pellet fraction, occurring as insoluble inclusion bodies.

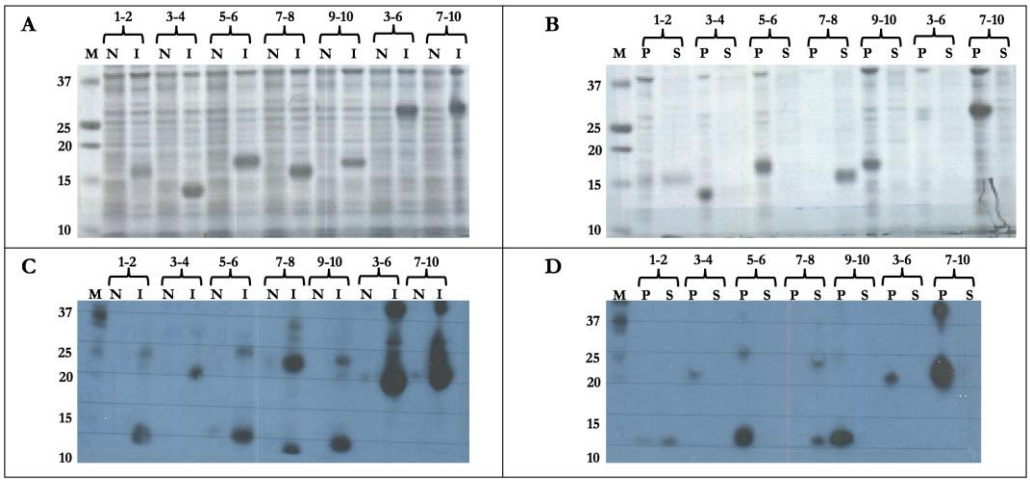


Figure 4.2. SDS-PAGE and Western Blot Analysis of PfRipr EGF Proteins: A. SDS-PAGE gel of NI and I fractions for each protein (15 µl of each sample loaded). B. SDS-PAGE gel of pellet and supernatant fractions for each protein. C. Western blot of NI and I fractions for each protein (1.5 µl of each sample loaded). The exposure time was three-seconds. D. Western blot of pellet and supernatant fractions for each protein. The exposure time was three-seconds. Abbreviations: M- molecular weight marker (kDa), N- Non-induced, I- Induced, P- Pellet, S- Supernatant.

Next, a scaled-up production of each protein was carried out. Induction with 1 mM IPTG following a temperature drop from 37 °C to 18 °C shifted most of the protein from the pellet fractions and into the soluble supernatant fraction (data not shown). Initial Ni-affinity chromatography was followed up by protein concentration determination for each 2 ml elution fraction then analysed on SDS-PAGE under reducing conditions (figure 4.3). Ni-affinity chromatography produced mostly monomeric protein; dimer present under reducing conditions is indicative of covalently-bonded monomers, and thus likely misfolded protein. However, some contaminating protein and degradation products remained as seen on the SDS-PAGE gel and Western blots (data not shown).

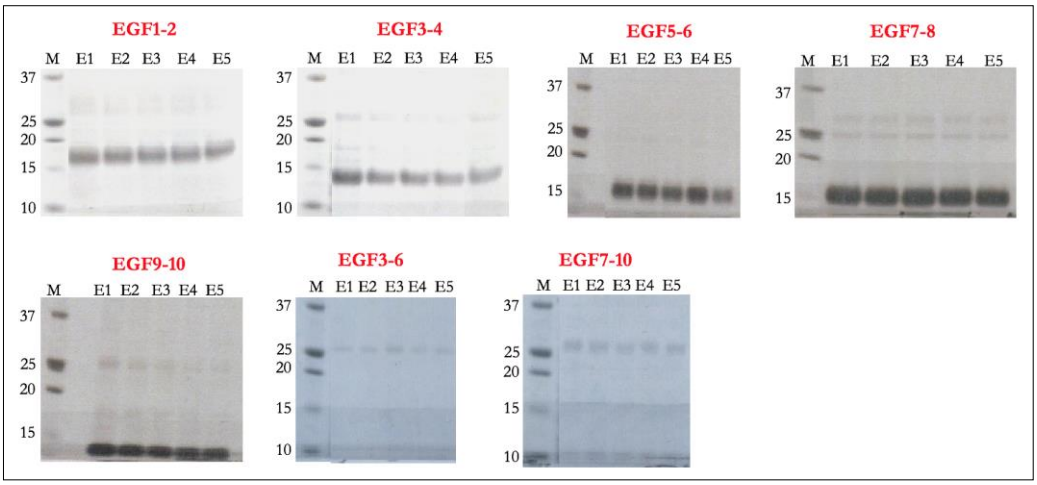


Figure 4.3. Ni-Affinity Purification of PfRipr EGF-like Domains: SDS-PAGE results of samples run under reducing conditions. EGF1-2 was produced from a 1.2 L culture and yielded the following concentrations of protein in each 4 ml fraction (following NanoDrop, all values in mg/ml): E1- 1.43, E2- 1.07, E3- 1.11, E4- 0.97, E5- 0.82. Around 3 µg of protein run for each fraction. EGF3-4 was produced from a 1.2 L culture and yielded the following mg/ml concentrations of protein in each 2 ml fraction: E1- 2.35, E2- 1.45, E3- 1.34, E4- 1.33, E5- 1.18. 2 µl of each sample was used to run on the gel. EGF5-6 was produced from a 1.2 L culture and yielded the following mg/ml concentrations of protein in each 2 ml fraction: E1- 0.71, E2- 0.62, E3- 0.65, E4- 0.63, E5- 0.62. 2 µl of each sample was used to run on the gel. EGF7-8 was produced from a 1.2 L culture and yielded the following mg/ml concentrations of protein in each 2 ml fraction: E1- 1.41, E2- 1.27, E3- 1.23, E4- 1.25, E5- 1.2. 2 µl of each sample was used to run on the gel. EGF9-10 was produced from a 300 ml culture and yielded the following mg/ml concentrations of protein in each 2 ml fraction: E1- 0.25, E2- 0.25, E3- 0.37, E4- 0.26, E5- 0.23. 2 µl of each sample was used to run on the gel. EGF3-6 was produced from a 600 ml culture and yielded the following mg/ml concentrations of protein in each 2 ml fraction: E1- -.4, E2- 0.17, E3- 0.16, E4- 0.16, E5- 0.16. 1 µg of each sample was used to run on the gel. EGF7-10 was produced from a 600 ml culture and yielded the following mg/ml concentrations of protein in each 2 ml fraction: E1- 0.37, E2- 0.23, E3- 0.23, E4- 0.23, E5- 0.24. 2 µl of each sample was used to run on the gel. Gel data shown are composites.

Each protein was concentrated to <5 ml and further purification carried out via SEC (data not shown). Various samples appeared to undergo significant sample loss after SEC, which was likely due to membrane-associated losses in both the Vivaspin® concentrators and 0.22 µl Merck filter units used prior to the

application of samples to the SEC column. Thereafter, instead of filtration, it was decided to remove particles by centrifuging samples at 10 000 × g for 20-minutes at 4 °C.

After concentrating key fractions to under 400 µl, they were run on SDS-PAGE (figure 4.4). As would normally be expected from disulfide-rich proteins, all EGF-like proteins ran faster under non-reducing conditions. Significant sample loss was seen in the flow-through of EGF3-4. This was due to accidentally using a 10-kDa molecular weight cut-off concentration tube rather than a 5-kDa one. It was also observed that EGF3-4 rapidly aggregated within a day of purification, with aggregation becoming worse over time (data not shown).

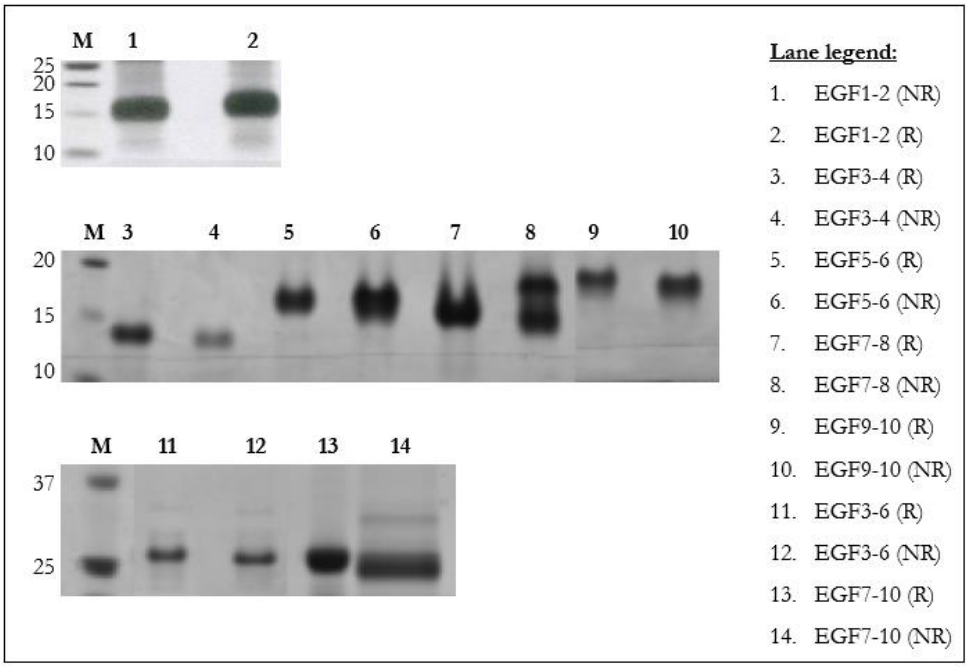


Figure 4.4. SDS-PAGE of PfRipr EGF-like Domains: None of the samples showed significant degradation under reducing or non-reducing conditions. Under NR conditions, EGF7-8 appears as two bands close together. EGFs 3-6 and 7-10 show the start of potential minor dimer formation. ~ 10 µg loaded for each sample. Gel data shown are composites.

In addition to the production of the EGF-like domains of PfRipr, the N-terminal (residues 21 – 287), mid-section (residues 326 – 636) and C-terminal (residues 978

– 1086) regions of the protein that were devoid of EGF domains were also of interest. The mid-section contains the region of PfRipr where there is enzymatic processing at an unknown site. Therefore, in addition to a regular N-terminal polyhistidine-tagged Ripr326-636 construct, an additional version was developed *in silico* containing the polyhistidine-tag at the C-terminus instead. This would allow for Edman degradation analysis after exposure to parasitic supernatant (containing the enzyme/s involved in PfRipr polypeptide processing) to identify the cleavage region. After the electroporation of the pME-T plasmids containing the synthetic genes into DH10β cells, >1000 colonies were seen on each LB agar plate. After a miniprep was carried out on selected colonies and the *Bam*HI/*Xho*I cut DNA isolated and purified (results not shown), ligation into pET45b+ plasmids were carried out. The resultant plasmids were used to transform BL21 gold cells resulting in thousands of colonies. Using selected colonies, test expression inducing at 30 °C with 0.25, 0.5 and 1-mM IPTG was performed followed by analysis of the results on SDS-PAGE (figure 4.5).

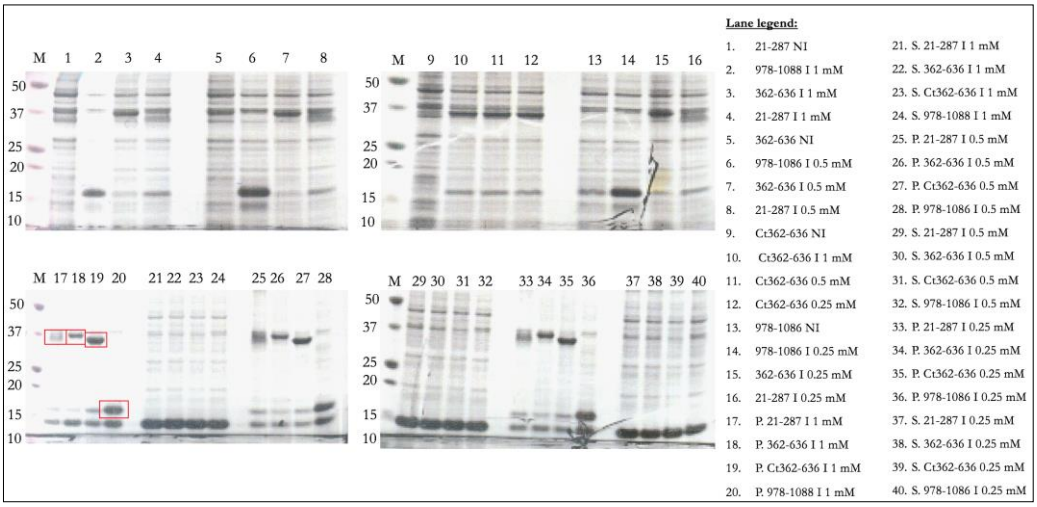


Figure 4.5. Test Expression of PfRipr Non-EGF Domain Proteins at 30 °C: Regardless of IPTG concentration, the majority of expressed protein appeared almost exclusively in the pellet fractions (red box highlighting the key bands for samples from 1 mM IPTG induction). Abbreviations: M- molecular weight marker (kDa), NI- non-induced sample, I- Induced sample (with concentration of IPTG given), P- pellet sample, S- supernatant sample.

Given that there was no discernible difference in protein expression regardless of IPTG concentration, an attempt to shift the destination of protein products was made by inducing with 0.25 mM IPTG at 18 °C. Samples were analysed on SDS-PAGE (figure 4.6). None of the sample appeared to express any detectible level of recombinant protein with the exception of Ripr21-287. However, in this case expression remained exclusively in the pellet fraction as insoluble inclusion bodies. To try and concentrate any protein present in the soluble fraction, the supernatant was applied to a Ni-NTA column, with the resulting elution fractions analysed via HPLC-ESI/MS (data not shown). However, this revealed no protein of interest. Refolding trials of the insoluble inclusion bodies from cultures induced at 30 °C were planned, however time limitations and the need to focus on the EGF-like domains prevented this from being carried out, in addition to planning for the later production of full-length PfRipr in either insect cells or *Pichia*.

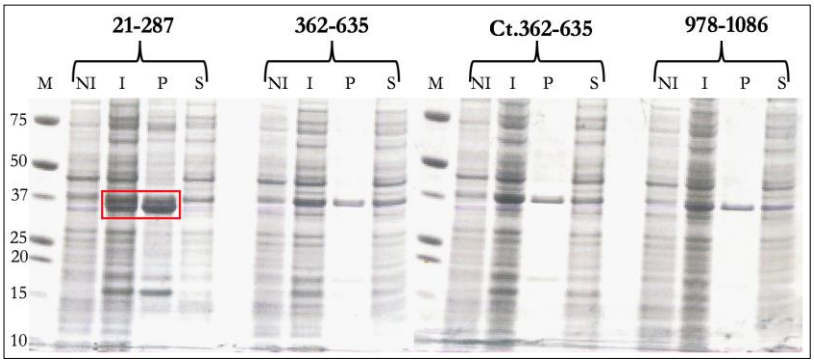


Figure 4.6. Test Expression of PfRipr Non-EGF Domain Proteins at 18 °C: No PfRipr non-EGF protein was shown to have any discernible protein expression aside from Ripr21-287 (red box). Ripr362-636 (and the Ct. version) may be masked by the contaminating *E. coli* protein around 37-kDa.

With respect to recombinant PfRh5, samples obtained from insect cell cultures were analysed by Western blot with anti-PfRh5 rabbit polyclonal antibodies (figure 4.7 A). All samples appeared to contain at least some PfRh5 expression. After harvesting, the supernatants were loaded onto an anti-flag affinity chromatography

column to purify the flag-tagged PfRh5, with no discernible protein loss in the flow-through or from protein concentration (figure 4.7 B).

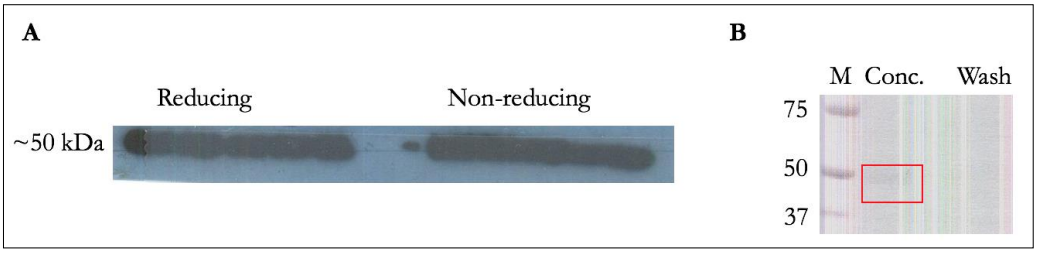


Figure 4.7. Western blot and SDS-PAGE of PfRh5: A. 20 µl samples were obtained from cell culture supernatant for each 200 ml flask and run on the gel under reducing or non-reducing conditions. The Western blot used 1:5000 anti-PfRh5 rabbit polyclonal antibodies and 1:2000 donkey anti-rabbit HRP. B. SDS-PAGE gel of post-anti-flag tag affinity chromatography concentrated sample (in red box) and the wash fraction from the concentration tube (showing no discernible sample loss).

The concentrated PfRh5 sample was further purified and buffer exchanged into 1 × PBS pH8 via SEC, and 2 µg of each fraction was analysed via SDS-PAGE (figure 4.8 A and B). Fractions 41 – 48 were combined and concentrated to 3 mg/ml (in 350 µl). 3.5 µl of the final concentrated PfRh5 protein was analysed on SDS-PAGE under reducing and non-reducing conditions to assess protein quality. The samples revealed no discernible protein aggregation and only minimal, insignificant degradation (figure 4.8 C). Under non-reducing conditions, PfRh5 runs slightly lower due to the presence of disulfide bonds. The production of purified PfRh5 and PfRipr proteins allowed for a variety of binding studies to be carried out, in addition to screening potential inhibitory compounds from the MMV box against PfRh5. It was also decided to explore an alternative protein production strategy for the PfRipr EGF-like modules, namely through refolding from inclusion bodies. This was, in part, due to significant batch-to-batch variation with the production of certain EGF-like proteins bacterially-produced.

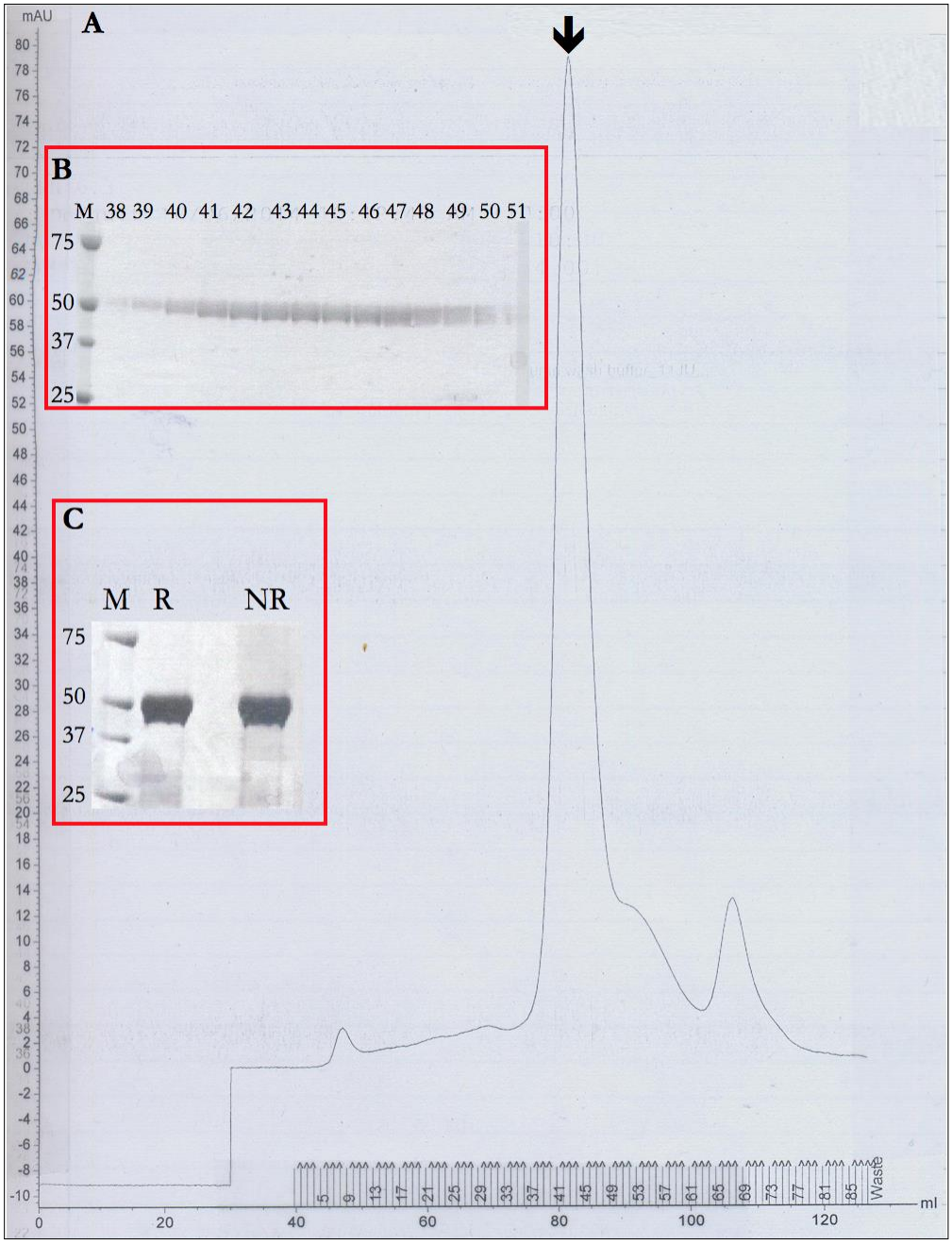


Figure 4.8. SEC and SDS-PAGE of PfRh5: A. On the SEC chromatogram, the arrow highlights the peak of interest. B. 2 µg from fractions 38 – 51 were analysed on an SDS-PAGE gel under reducing conditions. C. 3.5 µl of the final, concentrated product run under reducing and non-reducing conditions.

4.3 Ripr Protein Refolding from Inclusion Bodies

4.3.1 Methods for Ripr Protein Refolding from Inclusion Bodies

Recombinant protein expression, cell preparation and Ni-affinity chromatography of EGFs 5-6, 3-6 and 7-10 were carried out as described in sections 2.5, 2.6, and 2.7, respectively, with refolding as per the methods described in Haggarty-Weir, *et al.*, 2017.

The production of EGF3-10 was carried out as follows- Glycerol stocks of BL21 *E. coli* containing the EGF3-10 synthetic gene in a pET303 plasmid (glycerol stocks and general protein production/refolding methods provided by Xtal^{TM} Biostructures, USA) were streaked out on LB agar plates containing 40 μg/ml kanamycin and grown overnight at 37 °C. Colonies were picked into 4 ml of TB with 40 μg/ml kanamycin and incubated at 37 °C shaking at 250 rpm until the OD_{600} reached 0.5. Tube contents were added to 100 ml of fresh broth, grown as above, then 10 ml transferred into 500 ml of fresh broth. An aliquot of 1 mM IPTG was added when the OD_{600} reached 0.5, then the flask left to incubate for four-hours. Flask contents were centrifuged at 5000 × g for 15-minutes, supernatant decanted and 160 ml of lysis buffer (20 mM Tris pH 8, 200 mM NaCl, and four tablets of Roche cOmplete protease inhibitor) added. Sonication was carried out on each sample. The lysate was centrifuged at 17 000 × g for 45-minutes at 4 °C followed by separation of pellet and supernatant. The pellet was re-solubilized in 200 ml of IB wash buffer 1 (50 mM Tris pH 8, 250 mM NaCl, 1% Triton-X-100). After further sonication on ice the sample was re-centrifuged and the above two steps repeated. Upon the third time re-solubilizing the pellet, IB wash buffer 2 (20 mM Tris pH 8, 200 mM NaCl, and 1 mM $CaCl_2$) was used. The final pellet was solubilized in 5 ml/g of pellet of IB solubilisation buffer (6 M guanidine HCl, 20 mM Tris pH 8, 1% Triton-X-100, 20 mM βME, and 10 mM imidazole). Ni-affinity chromatography was carried out as described section 2.7,

but using a step-wise elution process with four elution buffers comprised of 6 M guanidine HCl, 20 mM Tris pH 8, 20 mM βME and increasing concentrations of imidazole (50, 100, 150, and 300 mM). SDS-PAGE was used to analyse fractions which were then concentrated to 8 ml as previously described.

Refolding of EGF3-10 commenced with adding 8 ml of protein solution in a slow, drop-wise manner into 400 ml of stirring refold buffer (50 mM Tris pH 10.5, 2 mM oxidized glutathione, 500 mM arginine, and 1 mM $CaCl_2$). The refold mix was incubated at 4 °C for 24-hours without stirring, then dialyzed against 8-L of dialysis buffer (20 mM Tris pH 8 and 100 mM NaCl) for 24-hours at 4 °C; this step was then repeated. Post-dialysis, the sample was concentrated to 40 ml, with 5 ml fractions used for SEC.

Trichloroacetic acid (TCA) precipitation was carried out on 1.1 ml of each sample. After 30-minutes centrifugation at 10 000 × g and 4 °C, the top 1.0 ml was removed and placed into tubes containing 0.1 ml 100% TCA. After a one-hour incubation on ice, samples were centrifuged for 15-minutes at 10 000 × g and 4 °C. Supernatant was decanted, 0.5 ml chilled acetone was added and then tubes re-centrifuged for five-minutes. After the supernatant was decanted, 20 µl of 1 x RSB or 1 x NRSB were added and samples analysed via SDS-PAGE (as described in section 2.3.1).

Dialysis was done using 3.5 kDa MWCO Spectra/Por™ 7 Membrane tubing (Fisher Scientific). Dialysis buffer was 4 x 5 L of 20 mM Tris pH8 and 20 mM NaCl. The dialysis was carried out on a magnetic stirrer plate at 4 °C, with buffer being changed twice a day for two days.

Dialysed proteins were centrifuged at 30 000 × g for 30-minutes before being pumped onto a 1 ml HiTrap Q HP anion-exchange column (GE Healthcare) that had been washed with elution buffer (20 mM Tris pH 8, 1 M NaCl) and equilibrated

with wash buffer (20 mM Tris, pH 8) The flow-through was re-pumped over the column before being attached to an ÄKTA pump system. A gradient of 0-100% elution buffer at a flow rate of 1 ml/minute was then used to elute the protein in 1 ml fractions. Key fractions were visualized via SDS-PAGE, combined as appropriate and concentrated to <5 ml final volume with a 5-kDa molecular weight cut-off Vivaspin® concentration tube (Satorius, Göttingen) via centrifugation at 4,200 × g for 15-minute intervals at 4°C.

The concentrated protein samples were then loaded onto an analytical Sigma 10/300 Sephadex 75 column (GE Healthcare) (S200 16/600 pg used for EGF7-10) that had been equilibrated with 1 × PBS pH 8 on an ÄKTA instrument (20 mM Tris pH 8 and 100 mM NaCl for EGF3-10). Protein was eluted at a rate of 0.8 ml/minute using 1× PBS pH 8 and collected in 0.5 or 1 ml fractions. The absorbance at 280 nM (A_{280}) of protein solutions during purification was used to calculate the minimum sample size required to load for SDS-PAGE analysis based on a visualization limit of 2 μg of protein on the gel. A_{280} was measured using the 'Protein' setting on an Eppendorf Biophotometer.

Western blot analysis was carried out as described in section 2.8, but using a 1:2000 dilution of rabbit anti-PfRipr polyclonal antibodies for the primary antibody (donated by Jenny Thompson at the Walter and Eliza Hall Institute; antibodies raised against PfRipr EGF3-10), and a 1:2000 dilution of anti-mouse HRP antibodies for the secondary.

4.3.2 Results of Ripr Protein Refolding from Inclusion Bodies

The solubilized inclusion bodies prepared from the initial 600 ml of S broth (induced with 1 mM IPTG at 37 °C) were applied to a Ni-NTA column, and the 23 x 1.5 ml elutions analysed via SDS-PAGE (figure 4.9). Due to the high level of

protein present following batch binding of the flow-through to Ni-NTA resin overnight, two more lots of elutions could be carried out over three days, with the fractions stored at -20 °C. The total amount of protein from the pooled fractions from the first elution was 59 mg for EGF5-6 and 55.36 mg for EGF3-6.

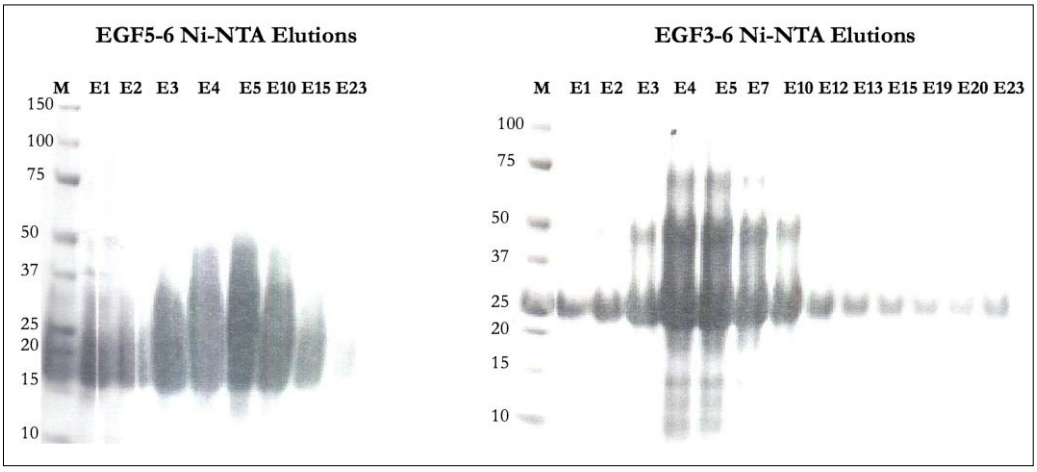


Figure 4.9. SDS-PAGE of Ni-Affinity Purification of EGF5-6 and EGF3-6: 20 µl loaded onto each lane. All fractions combined and stored at 4 °C until required.

An aliquot of 1.1 ml from each pooled elution was used for each test refold condition attempted (20 ml final volume). The refold trials were carried out over three days at room temperature, with 1 ml samples taken for TCA precipitation and subsequent SDS-PAGE analyses (figure 4.10). Initial samples (t_0) show the presence of numerous multimers for both proteins, which are largely not present in the t_{72} hour samples where the only significant band aside from monomer is dimer (EGF3-6 still had the presence of some trimer). All samples run with 1 x NRSB appear to migrate faster on the gel, indicating the presence of disulfide bond formation. There did not appear to be any significant difference between samples that underwent refolding in reducing conditions compared to those refolded in non-reducing conditions. Additionally, increasing urea concentration appeared to result in less protein visible on a gel compared to no urea.

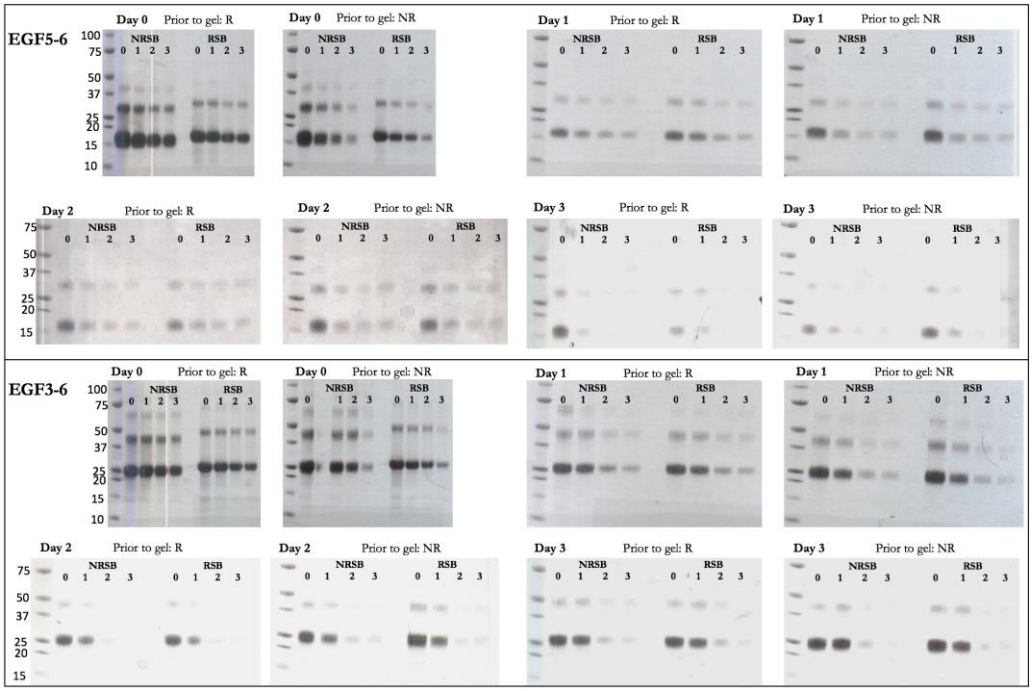


Figure 4.10. EGF5-6 and EGF3-6 Refolding Trials: Each sample is from a TCA precipitation carried out during the refolding period. Above each gel indicates whether the samples prior to SDS-PAGE were in R or NR refolding conditions. Numbers atop each gel lane indicate the amount of urea used (M). Above this are the conditions the samples were run under; RSB or NRSB. Time is in days in bold in the top left-hand corner of each gel.

Next, it was decided to repeat the refolding trial, but using different amounts of oxidized glutathione (1, 0.5 or 0.25 mM) to assess its effect on refolding. The results for t_0 and t_{72} samples are shown in figure 4.11. For both proteins, it appears that using 0.25 mM oxidized glutathione was perfectly sufficient as judged by there being no significant difference in multimer formation over time.

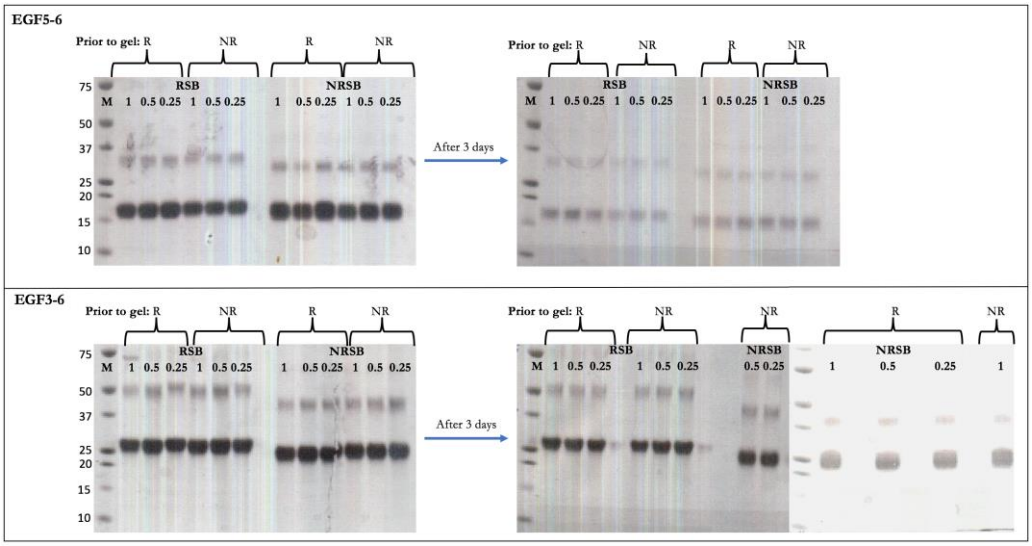


Figure 4.11. EGF5-6 and EGF3-6 Refolding Trials with Oxidized Glutathione: Each sample is from TCA precipitations carried out during the refolding period (day 0 and day 3). Above each gel is indicated whether the samples prior to SDS-PAGE were in R or NR refolding conditions. Numbers atop each gel lane indicate the amount of oxidized glutathione used (M).

Subsequently, scaled-up production of refolded EGFs 5-6, 3-6, 7-10 and 3-10 (it should be noted that the 34.5 ml stock of EGF7-10 contained 49 mg of protein, and for the 8 ml of EGF3-10 there was a total of 16 mg) was carried out using 10 ml of the stock protein solution. After refolding and two days of dialysis to reduce the salt content to 20 mM, the samples were analysed by SDS-PAGE (figure 4.12). The dialyzed samples do not show any significant loss of protein, but do show some slightly stronger trimer and tetramer bands than had been previously been seen. This may be the result of a lower salt concentration, namely going from 100 to 20 mM NaCl.

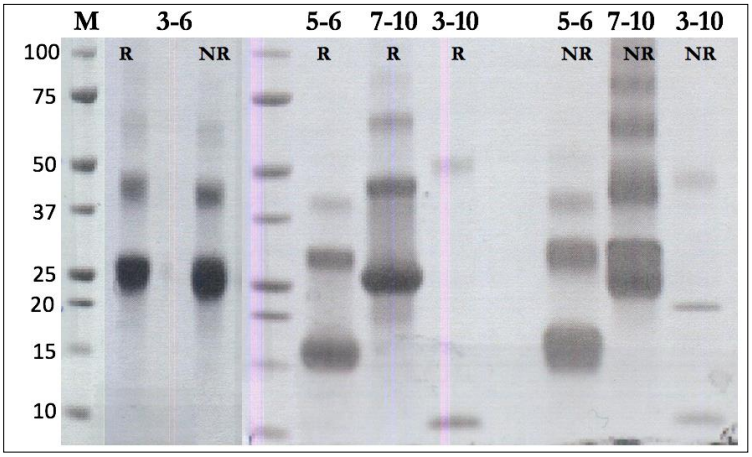


Figure 4.12. Scale-up Refold of EGFs 5-6 and 3-6, 7-10, and 3-10. Each sample is from TCA precipitations carried out (20 µl loaded onto the gel. Sample running conditions (R or NR) for the SDS-PAGE is indicated at the top of each sample gel lane.

Anion-exchange chromatography was carried out on each protein (with the exception of EGF3-10) to remove the higher order multimers. The results in figures 4.13 to 4.15 show that the multimers, with the exception of dimer, could be separated from monomer. Fractions 23 and 24 from EGF5-6 (figure 4.13) were combined and concentrated down to 0.5 ml. Fractions 25 – 28 from EGF3-6 (figure 4.14) were combined and concentrated down to 0.5 ml. Fractions 8 – 13 from EGF7-10 (figure 4.14) were combined and concentrated down to 0.5 ml.

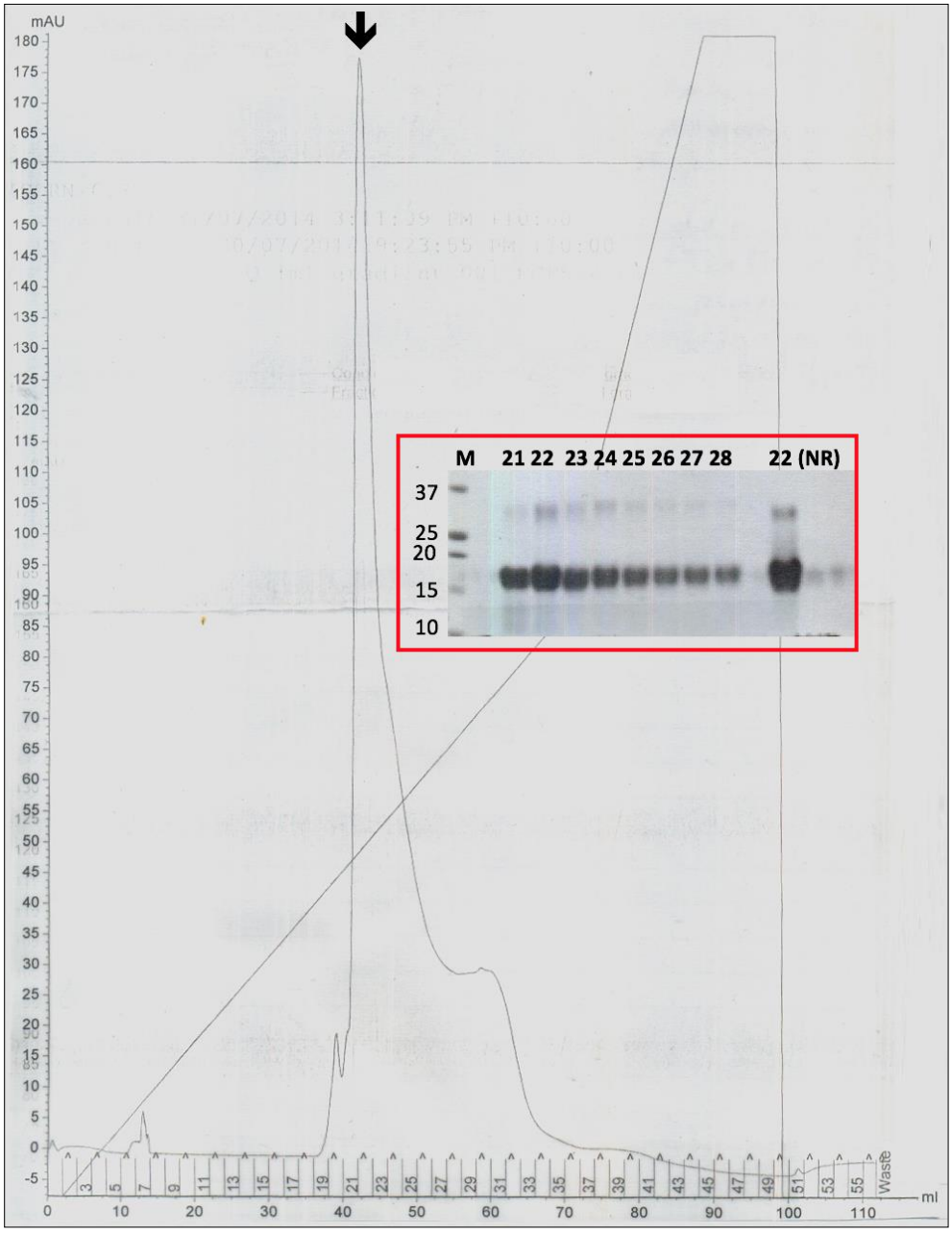


Figure 4.13. Anion Exchange Chromatography of EGF5-6: Protein elution occurred around 25% buffer B. Arrow indicates the main peak of interest. 2 ml fractions collected. Around 4 µg loaded onto the SDS-PAGE gel with samples run with 1 x RSB (unless otherwise noted with NR).

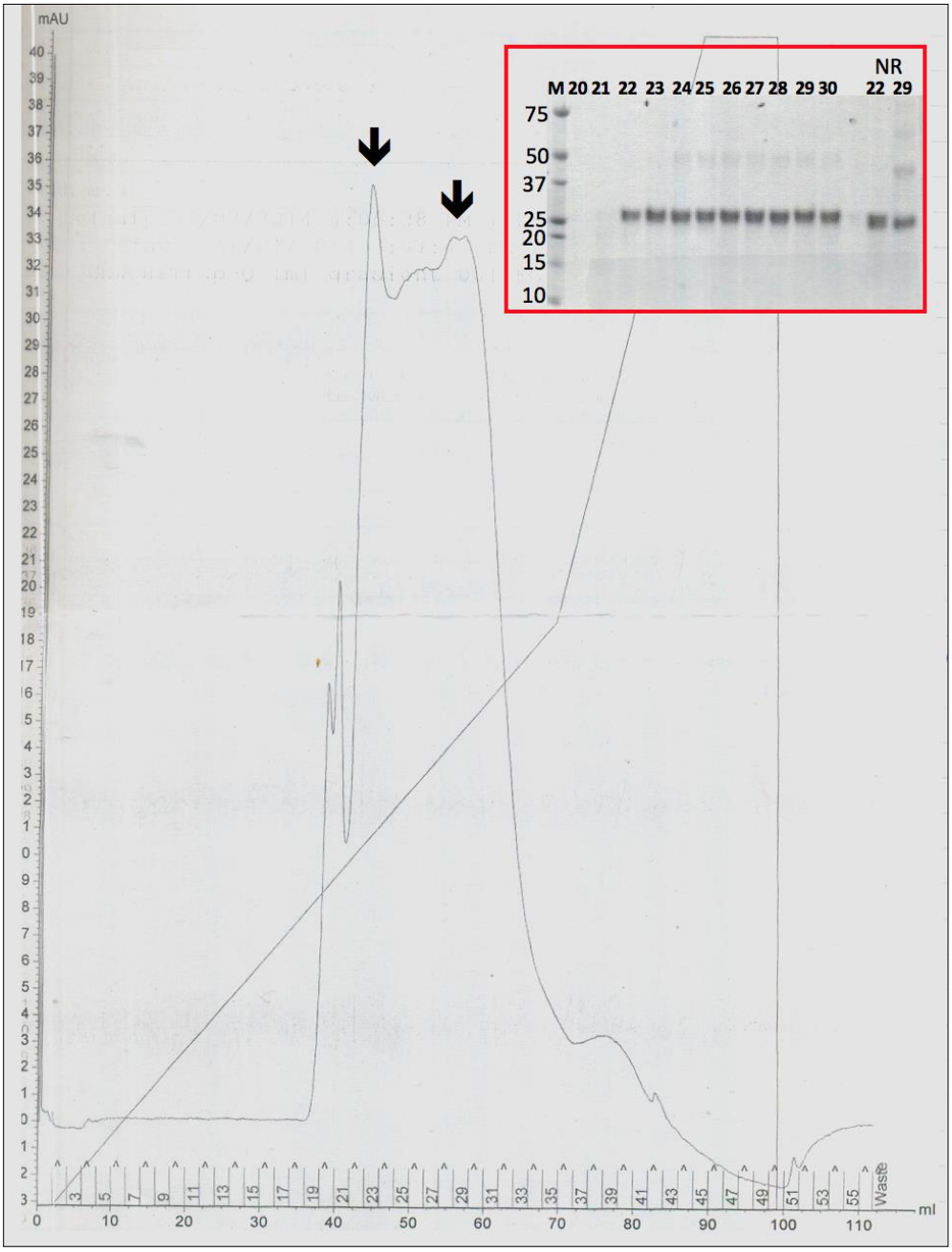


Figure 4.14. Anion Exchange Chromatography of EGF3-6: Protein elution occurred around 25% buffer B. Arrows indicate the main peaks of interest. 2 ml fractions collected. Around 4 µg loaded onto the SDS-PAGE gel with samples run with 1 x RSB (unless otherwise noted with NR).

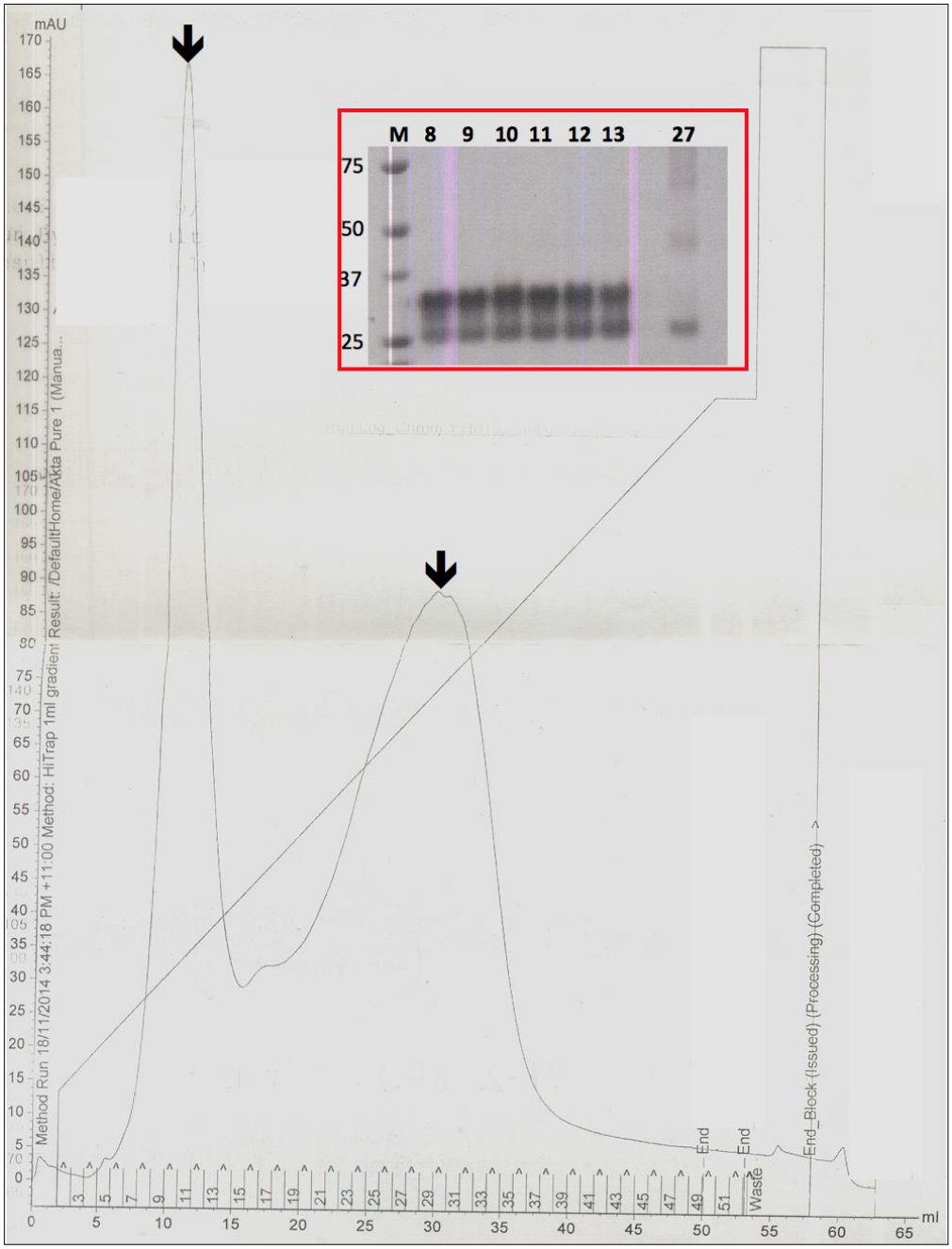


Figure 4.15. Anion Exchange Chromatography of EGF7-10: Protein elution occurred around 15% buffer B. Arrow indicates the main peaks of interest, with the initial, taller and sharper peak containing the majority of monomeric protein (the second, broader peak contains a variety of multimers). 1 ml fractions collected. Around 4 µg loaded onto the SDS-PAGE gel with 1 x NRSB.

Once all samples were concentrated to appropriate volumes, and after being centrifuged at 10 000 × g for ten-minutes at 4 °C, they underwent SEC to try to

separate the dimers from the monomers. SDS-PAGE analysis was used to visualize the key fractions, followed by a Western blot analysis to ensure each refolded protein was recognized by anti-PfRipr antibodies. The resultant chromatograms, SDS-PAGE gels and Western blot images are shown in figures 4.16 – 4.19. A slight amount of proteolytically clipped EGF3-6 remained, as did a minor proportion of EGF7-10 dimer. EGF5-6 appeared purely monomeric. All proteins were recognized by the anti-PfRipr antibodies as expected.

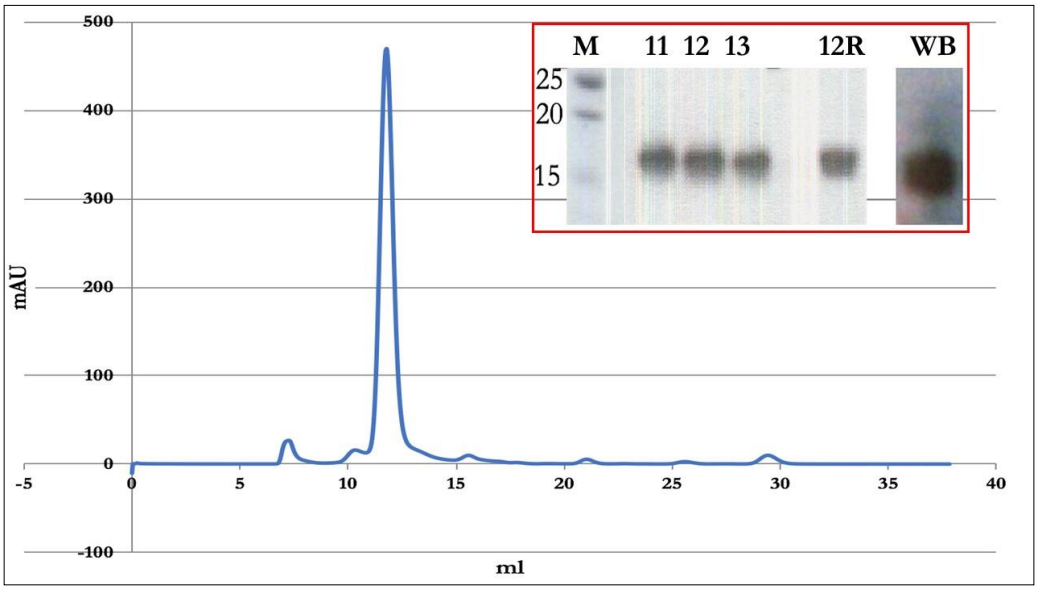


Figure 4.16. SEC, SDS-PAGE and Western Blot of EGF5-6: One major peak shown on the SEC chromatogram. In each case, 2 µg was analysed on SDS-PAGE under non-reducing conditions (centre fraction also analysed with 1 x RSB). Western blot (WB) carried out on 0.1 µg of protein from fraction 12 run with 1 x NRSB with a one-second development time.

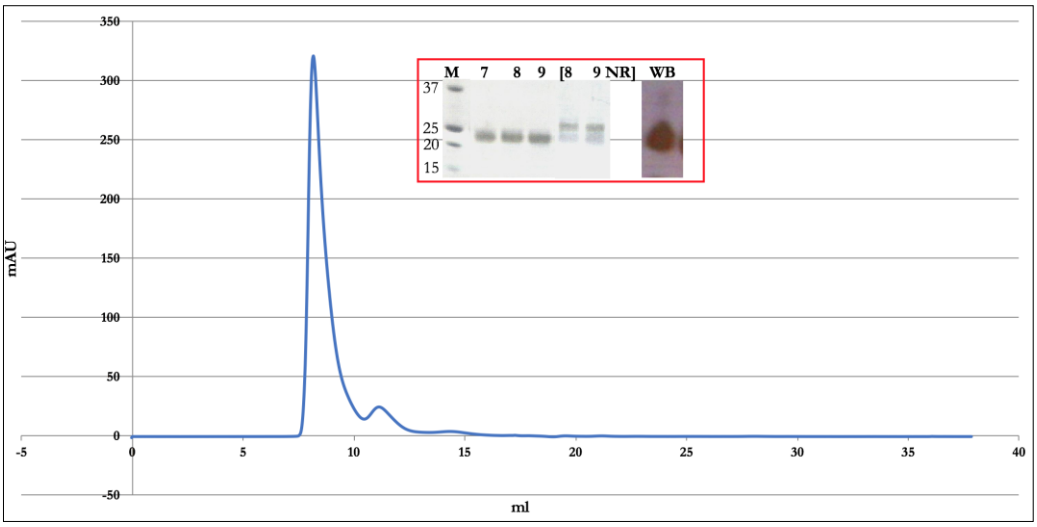


Figure 4.17. SEC, SDS-PAGE and Western Blot of EGF3-6: Three major peaks shown on the SEC chromatogram, with the major tall peak indicating the monomer. In each case, 2 µg of protein were analysed on the SDS-PAGE gel, under reducing conditions (fraction 8 and 9 also analysed with 1 x NRSB). Western blot (WB) carried out on 0.2 µg of protein from fraction 9 run with 1 x NRSB with a one-second development time.

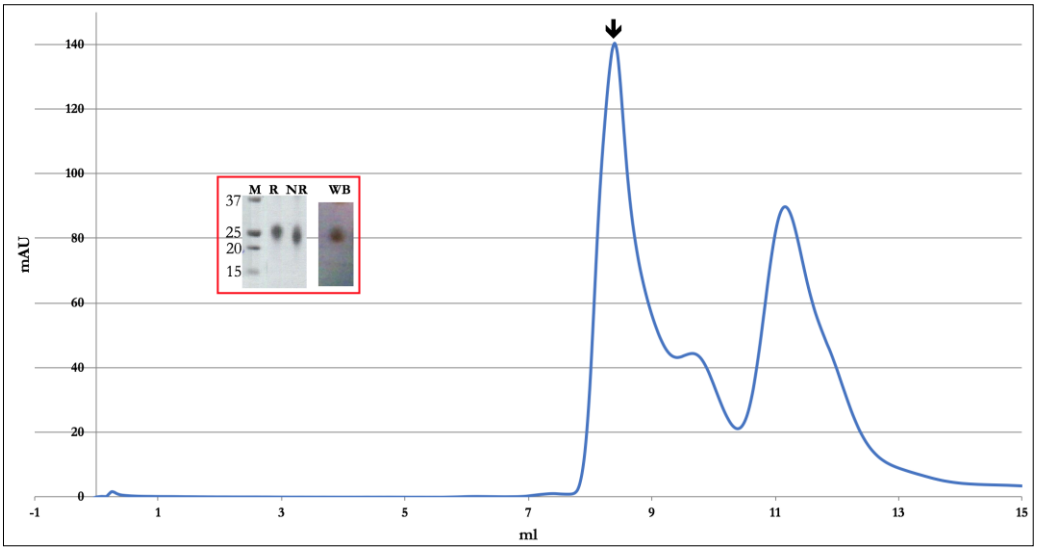


Figure 4.18. SEC, SDS-PAGE and Western Blot of EGF7-10: Samples were taken across the large peak shown (arrow), with 2 µg of fraction 8 analysed on an SDS-PAGE gel under R and NR conditions. Western blot (WB) carried out on 0.2 µg of protein from fraction 8 run with 1 x NRSB with a one-second development time.

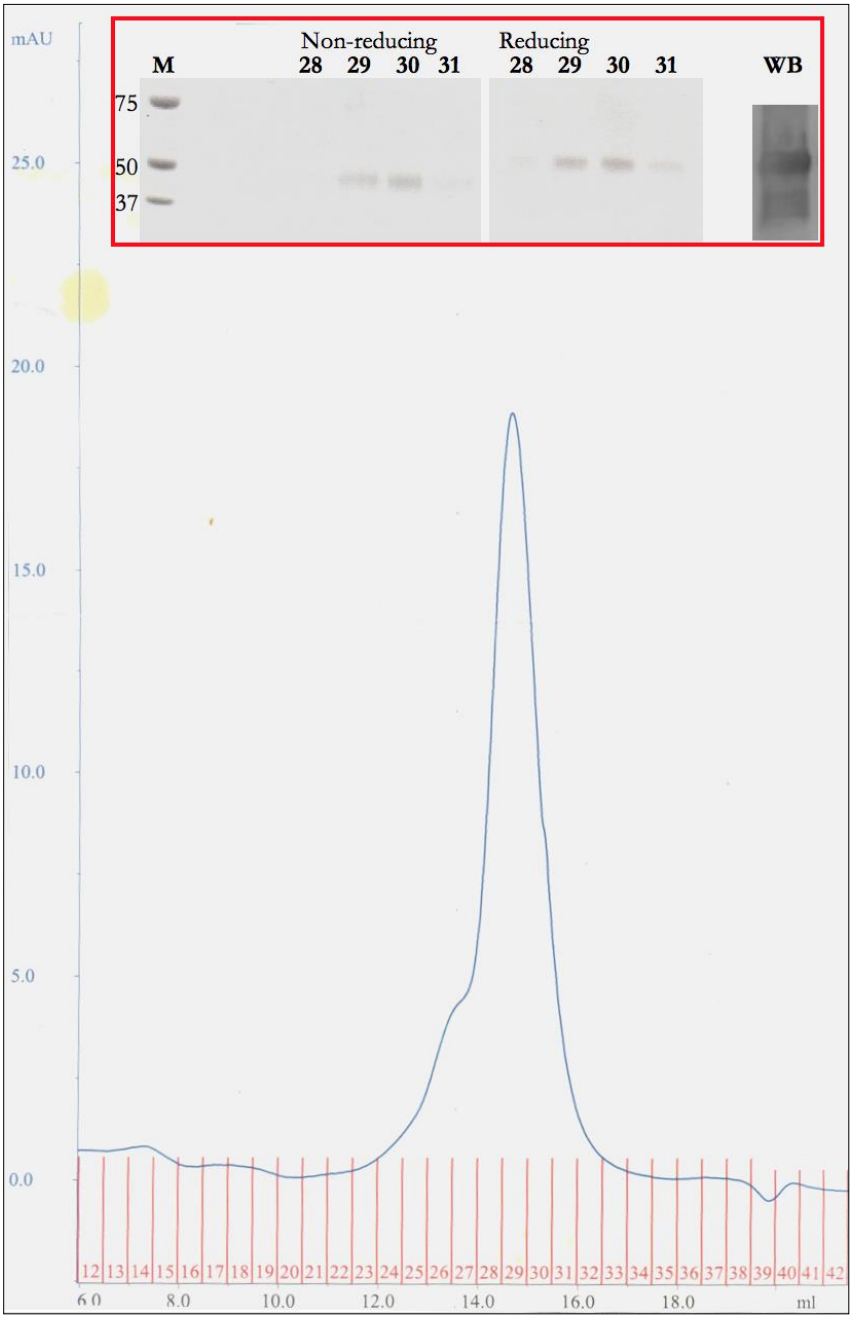


Figure 4.19. SEC, SDS-PAGE and Western Blot of EGF3-10: One major peak with a minor shoulder is shown in the chromatogram. Samples were taken to run under NR and R conditions on SDS-PAGE from under the centre of the major peak. The Western blot (WB) shown here is actually from a prior SEC run corresponding to ~1 µg of fraction 28 here, run with 1 x RSB.

Finally, key fractions of monomeric protein were taken and concentrated to under 0.3 ml for a final concentration of ~12 mg/ml. This however excluded EGF3-10,

which could not be concentrated above 3 mg/ml without precipitating out of solution (samples remained at 0.5 mg/ml). Were there was insufficient quantity or purity for future experiments, more protein was made with the above protocol, but using a 1:25 dilution into the refold buffer rather than 1:100 (same results obtained, data not shown). This ensured pure monomeric protein was used for future experiments (such as crystallography), with the exception of EGF3-10, which (regardless of concentration) always appeared to reform a minor amount of dimer; around 5-10% of total protein concentration. After the final purification, each protein sample needed to be assessed for correct folding via biophysical methods. This was especially important as, whilst each protein could be detected by anti-PfRipr antibodies, it is not necessarily indicative of correct folding due to the antibodies being polyclonal and targeting both linear and structural epitopes. To first ensure sample quality, mass spectrometry was carried out followed by CD and NMR to assess the secondary structure and folding of each protein.

4.4 Biophysical Characterization of EGF-Rich Modules

4.4.1 Methods for the Biophysical Characterization of EGF-Rich Modules

4.4.1.1 Mass Spectrometry

Mass spectrometry was carried out at the following facilities: Bio21 Mass Spectrometry facility (Melbourne, Australia), the Scottish Instrumentation and Resource Centre for Advanced Mass Spectrometry (SIRCAMS) at the University of Edinburgh School of Chemistry (Scotland, United Kingdom), and the BSRC mass spectrometry and proteomics facility of St. Andrews University (Scotland, United Kingdom).

Matrix Assisted Laser Desorption Ionization-Time of Flight (MALDI-TOF) spectra were collected on a Daltonics ultrafleXtreme™ MALDI-TOF (Bruker) operating in

linear positive mode. Protein samples were prepared 1:1 with synapinic acid (50:50 acetonitrile (ACN):H_2O, 0.1% TFA. The lowest laser intensity that gave a clean peak was used.

Positive Electrospray Ionisation mass spectra (ESI+/MS) were collected on two instruments both equipped with analytical HPLC to identify peaks corresponding to the monoisotopic masses predicted for each intact peptide:

i) ESI Waters SYNAPT G2 Q-TOF equipped with a SYNAPT™ Acquity UPLC using a 1 mm x 250 mm proswift RP4H (PSDVB) column and a gradient of 5% A to 60% B (A= 0.1% formic acid (FA) in H_2O, B = 0.1% FA in ACN).

ii) 12T Bruker Daltonics solariX FT-ICR equipped with a Dionex™ U3000 HPLC using a 500 µm x 100 mm PROSWIFT L4 RP5H column and gradient identical to that in i).

4.4.1.2 Circular Dichroism

CD measurements were carried out on a JASCO J-180 spectropolarimeter (initial analysis at the Bio21 Institute of the University of Melbourne, and further analysis carried out at the Edinburgh Protein Production Facility (EPPF) of the University of Edinburgh by Roma Galloway with assistance from Dr. Elizabeth Blackburn). Far-UV CD spectra (185–260 nm) were collected using a 14 µM protein concentration in 10 mM $NaPO_4$ and 100 mM NaF and a 0.2-mm path-length cell. A bandwidth of 1 nm was used with a scan speed of 10 nm/min. All CD measurements were made at 25°C, and each spectrum was the average of five individual scans. The spectra were corrected for buffer contribution. DichroWeb software (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) was used for the quantitative analysis of CD spectra (Whitmore & Wallace, 2004).

4.4.1.3 Thermal Shift Assay

The fluorescence-based thermal shift assay (TSA) was carried out at the Walter and Eliza Hall Institute with the assistance of Dr. Isabelle Lucet according to the general methods outlined in (Lucet, *et al.*, 2014).

4.4.1.4 PfRipr PEGylation

For these experiments, 3 μg of each protein was prepared in duplicate; half of these were boiled for one-minute at 95 °C. Then 0.5 mM of methoxypolyethylene glycol malemide (10, 000 MW average, Sigma Aldrich) was added to each sample followed by one-hour incubation at room temperature. The reaction was quenched with 200 mM DTT, then the samples analysed alongside controls (3 μg of protein with no additional reagents) on SDS-PAGE with 1 x RSB.

4.4.1.5 Nuclear Magnetic Resonance Spectroscopy

NMR experiments were performed on an AVANCE 800 MHz (Bruker) spectrometer using a 5-mm triple resonance cryoprobe. Protein samples were between 150 μM - 300 μM in 10% D2O. NMR data was acquired using Bruker TOPSPIN 1.2 software and analysed using MestReNova software. The NMR experiments on EGF5-6 that was produced from a pProExHTA plasmid were carried out alongside Roma Galloway (refer to Galloway, 2016). Juraj Bella, Dr. Brian Smith, A/Prof. Matthew Call and Profs. Paul Barlow and Dusan Uhrin provided advice on experimental set up. Juraj Bella carried out the Bruker-based experimental set up and pulse sequence programming.

Production of ^{15}N isotopically labelled protein was carried out using *E. coli* grown in

minimal media consisting of (per L): 100 µl 1 M $CaCl_2$, 1 ml 2 M $MgSO_4$, 5.96 g anhydrous Na2HPO_4 dibasic, 3 g KH_2PO_4 monobasic, 1 g ^{15}N-NH_4Cl, 0.5 g NaCl, 4 g D-glucose, 4 ml of a Centrum A-Zinc solution (stock: one tablet dissolved in 40 ml bi-distilled H_2O, centrifuged at 5000 × g for 20-minutes then sterile filtered with a 0.22 µm filter unit); used to provide trace metals.

4.4.2 Results: Biophysical Characterization of EGF-Rich Modules

4.4.2.1 Mass Spectrometry

Mass spectrometry was carried out on all refolded protein samples to ensure correct mass identification prior to further biophysical characterization. Figure 4.20 shows the results for each refolded protein, in addition to EGF1-2 from figure 4.4 (a protein produced via soluble expression in *E. coli*). Each protein was identified as having the correct molecular mass, with a slightly higher mass observed for EGF3-10; around 0.48 kDa higher. This may in part be due to the presence of a cloning artefact from the pET303/CtHis vector, Arg-Ser-Arg at the N-terminus, though this would only account for 0.453 kDa, leaving 27 Da unaccounted for.

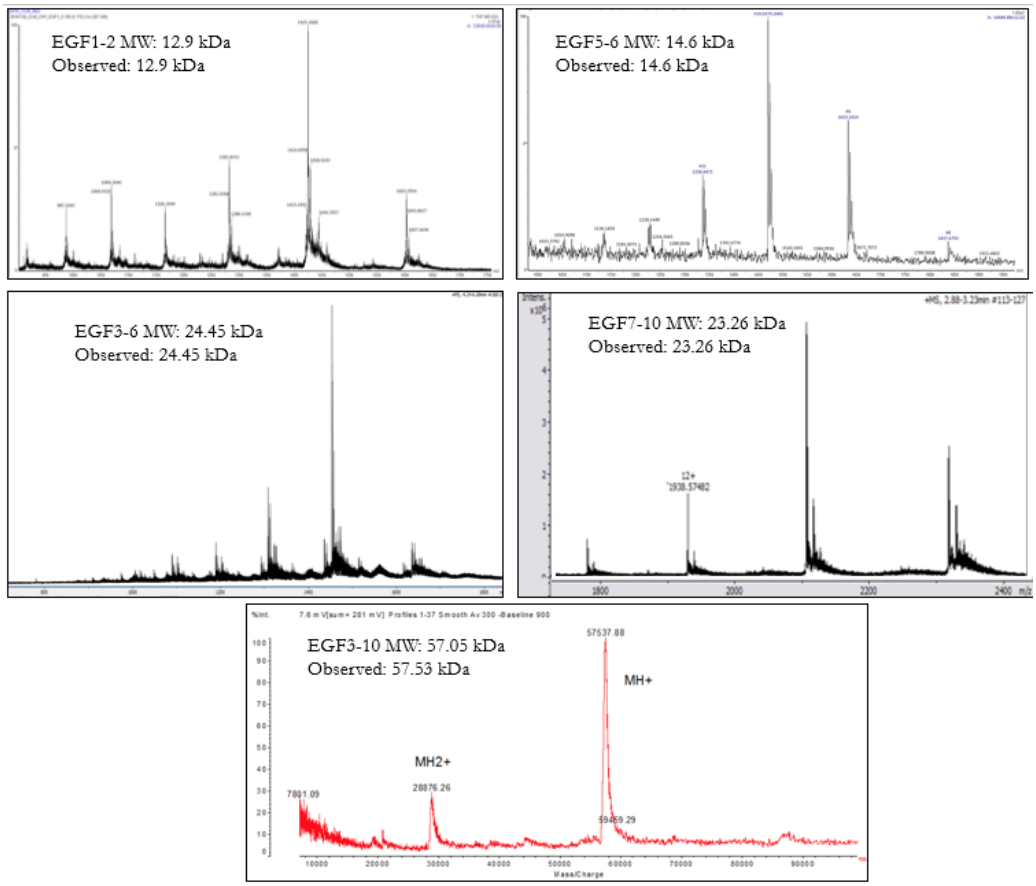


Figure 4.20. Mass Spectrometry Analysis of PfRipr EGF Proteins: Correct mass identification of each EGF protein was carried out via LC/ESI-MS, with the exception of EGF3-10 which was analysed via MALDI-TOF. Predicted and observed masses given for each sample.

4.4.2.2 Circular Dichroism

The amino acids that contribute to the antiparallel β-strand found in EGF domains typically account for approximately 17-22% of the primary sequence. The PfRipr EGF-like protein truncations produced in this chapter contain tandem EGF-like modules, in addition to undefined N- and C-terminal sequences, and in the case of EGF5-6 and EGF3-6, a predicted unusual long linker of 17 amino acid residues between the 5th and 6th EGF-like modules. CD was employed to analyse refolded EGF proteins for both the presence of β-strand (suggestive of correct protein folding), and to assess if any other secondary structure was present. CD

analysis of EGF3-10 was not carried out at this point in the project, due to the decision to attempt production of this protein in a *Pichia* expression system (where refolding would be unnecessary). Due to equipment repairs carried out on the Bio21 Institute CD machine, refolded EGF7-10 could not be analysed (additionally, this protein was going to be moved to a *Pichia* expression system). The initial CD analysis of EGF5-6 and EGF3-6 is shown in figure 4.21.

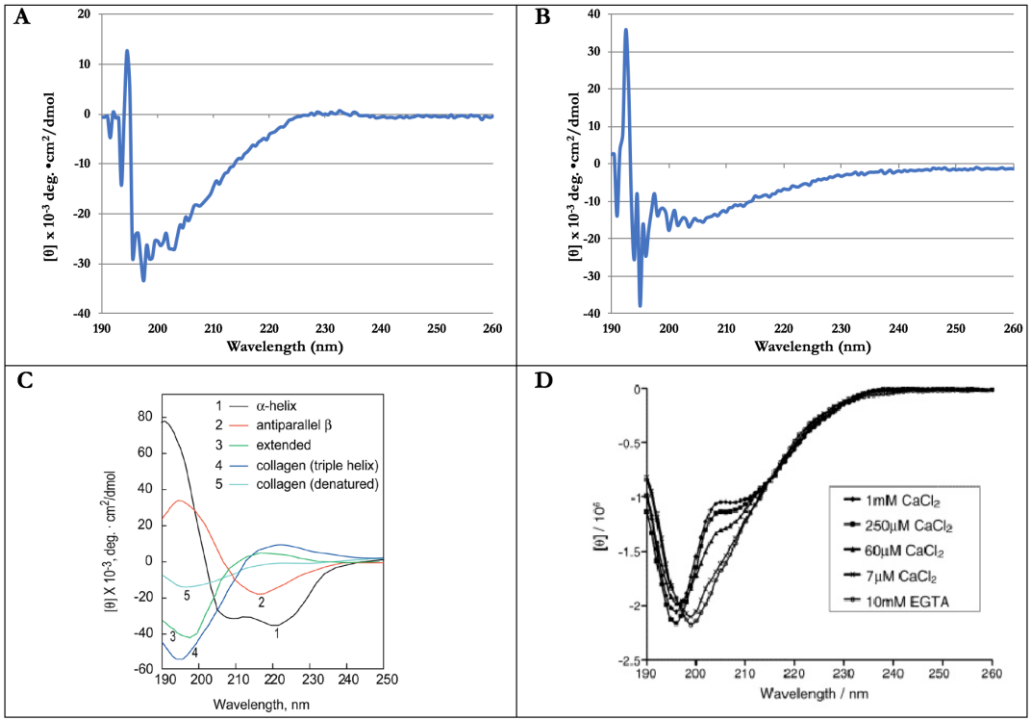


Figure 4.21. Initial CD Analysis of EGFs 5-6 and 3-6: A. CD of 14 μM EGF5-6 in a buffer comprised of 20 mM Tris pH 8, 100 mM NaF (plus some Roche protease inhibitor). B. CD of 10 μM EGF3-6 in the same buffer. C. An example of various secondary structure CD spectra traces, in addition to examples of folded and denatured proteins (adapted from (Greenfield, 2006)). D. Example of a CD spectra obtained from refolded EGF modules of a *P. falciparum* micronemal protein, cbEGF7-9 (Periz, *et al.*, 2005). CD spectra for both EGF5-6 and 3-6 were corrected by subtraction of spectra acquired for the buffer-only.

The CD spectrum of EGF5-6 and EGF3-6 displays minimums of ~198 nm and ~195 (respectively), which is characteristic of a protein in a predominantly random coil structure. No evidence for α-helix was shown for either protein, given the

absence of a double-dip starting around 220 nm and ending around 205 nm, in addition to the paucity of highly positive ellipticity at 190 nm. A positive peak is seen 195 and 193 nm for EGF5-6 and EGF3-6, respectively, with both dipping back down towards the furthest end of the UV spectrum. This is similar to the general peak shape shown for the β-sheet-containing protein in figure 3.30 C. However, the jagged peak structure towards this far UV end of the CD spectrum is also likely to have had a significant contribution from the sample buffer. It was also obvious that neither protein was predominantly β-sheet due to a lack of peak minima around 218 nm, which was expected.

A second sample of 14 μM EGF5-6 was prepared in 10 mM $NaPO_4$ + 100 mM NaF (no protease inhibitors) and re-analysed via CD (figure 4.22). Secondary structure content for EGF5-6 was estimated using the CONTIN deconvolution algorithm on the DichroWeb server (Whitmore & Wallace, 2004). This revealed the protein to be around 60% disordered, ~22% β-strand, with ~15% comprised of turns (plus a negligible amount of potential helical content). The peak minima remained at ~198 nm. This was in-line with the triple-EGF repeat protein CD shown in figure 3.30 D. The "goodness of fit" parameter NRMSD (normalised root mean square deviation) of EGF5-6 in figure 4.22 was reported as being 0.043. This is the most useful single measure of how well the theoretical CD spectrum calculated from the derived secondary structure composition matches the given experimental data over the wavelength range recorded (Kelly, Jess, & Price, 2005). NRMSD ranges from 0 (perfect fit) to 1 (no fit), with values lower than 0.1 (ideally lower than 0.05) to be deemed reliable (Kelly, Jess, & Price, 2005). Given this, the CD spectra for EGF5-6 should be regarded as high quality in terms of how accurate the data matches the theoretical information.

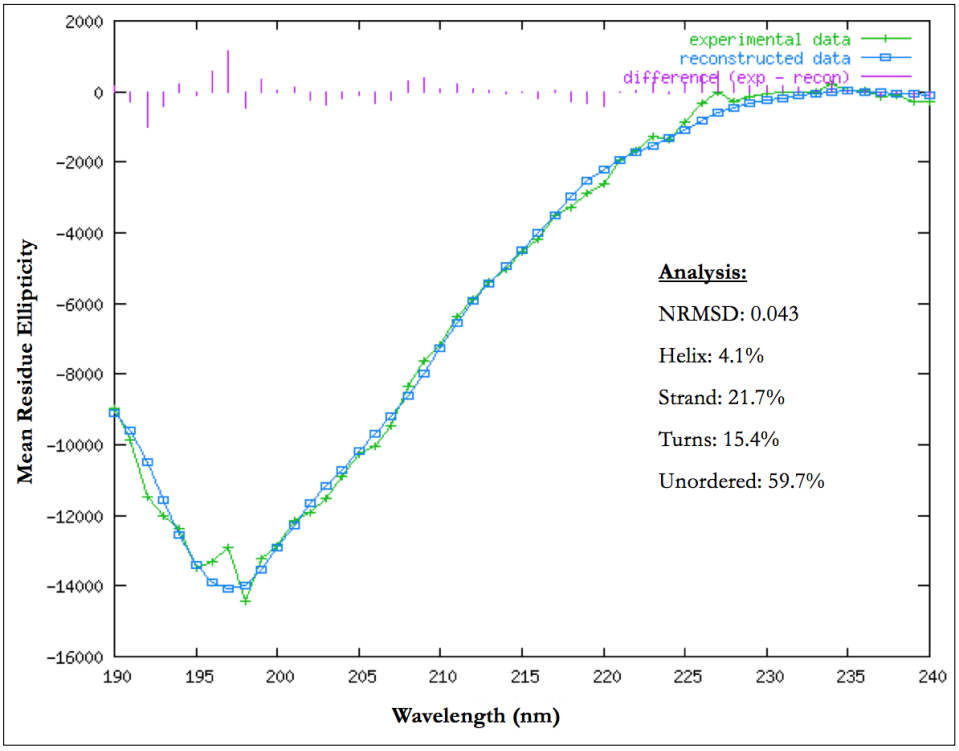


Figure 4.22. Far UV CD Spectra of EGF5-6: The DichroWeb-generated CD spectrum for a 14 μM EGF5-6 (10 mM $NaPO_4$ + 100 mM NaF) sample including quantitative analysis information. Residuals of the fit to solutions from a reference data set (Sreerama & Woody, 2000) (pink) indicate a good agreement between experimental and reconstructed data.

4.4.2.3 Thermal Shift Assay

To assess both the folding and thermal stability of the refolded EGFs 5-6 and 3-6, a TSA using the fluorophore CYPRO Orange that binds to hydrophobic patches on proteins was employed (figure 4.23). Both proteins were judged to be folded and highly stable, with little binding of the dye (i.e. exposure of hydrophobic residues) below 65 °C. The EGF3-6 sample had a slightly higher melting temperature than the EGF5-6 sample (78 °C for EGF5-6 and 79.5 °C for EGF3-6), though both displayed a similar profile. A side hump on the melting curve was seen at around 74 °C for both samples. This could be due to differential melting temperatures for the component domains or possibly due to a proportion of dimer

in the sample. However, overall, the TSA revealed the proteins to have evidence of being folded; unfolded protein would take up the dye faster at an earlier time point leading to a saturation of fluorescent signal.

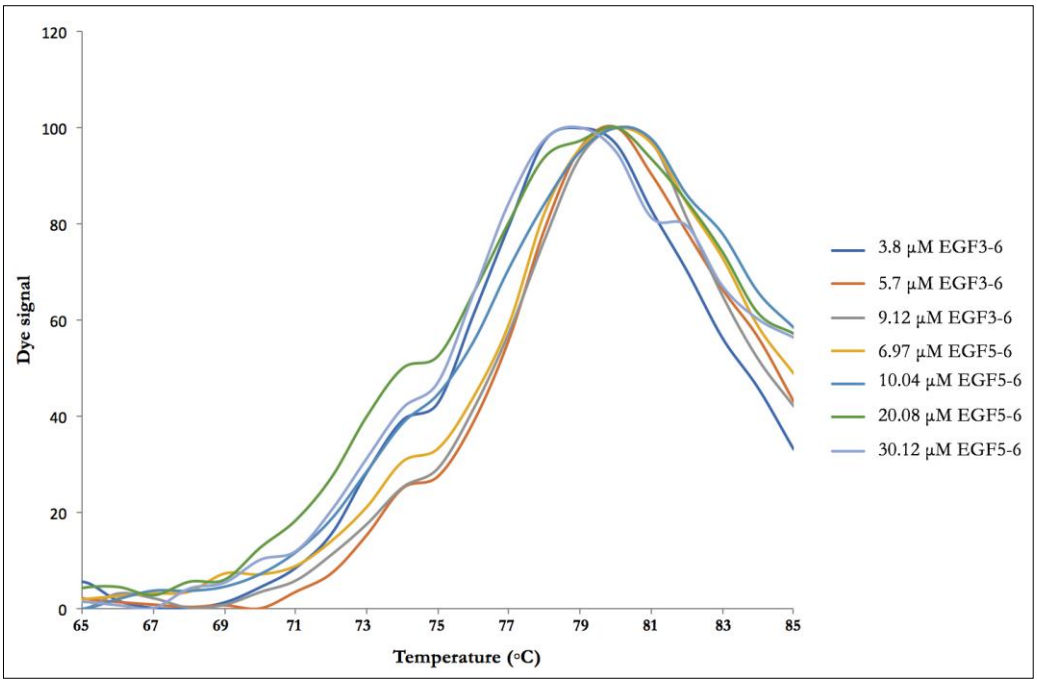


Figure 4.23. TSA of EGFs 5-6 and 3-6: Increasing concentrations of both EGF5-6 and EGF3-6 were used for a comparative analysis. Temperature raised 1 °C per 0.5-minutes.

4.4.2.4 PfRipr PEGylation

In order to assess the presence of free cysteine residues, a PEGylation reaction was carried out using methoxypolyethylene glycol malemide that would increase the molecular weight of each protein by 10 kDa per free cysteine bound. Refolded EGFs 5-6, 7-10 and 3-10 were analysed via SDS-PAGE (figure 4.24). Regarding the control samples (no PEG), both EGFs 5-7 and 7-10 appeared as monomers (as expected), whilst EGF3-10 showed a small percentage of dimer at around 100 kDa (as had been seen previously). With respect to the experimental samples, denaturation via boiling had no observable effect on the extent of PEGylation,

which manifested as faint bands running around 10 kDa higher than the respective non-PEGylated proteins. All proteins contain an even number of cysteines that are predicted to be forming disulfide bonds. These results suggest that only a small proportion of the protein (probably <5%) has a free cysteine following refolding, which is not uncommon. This data supports the case for successful refolding of more than 95% of the protein in the sample, which is sufficient for most analyses.

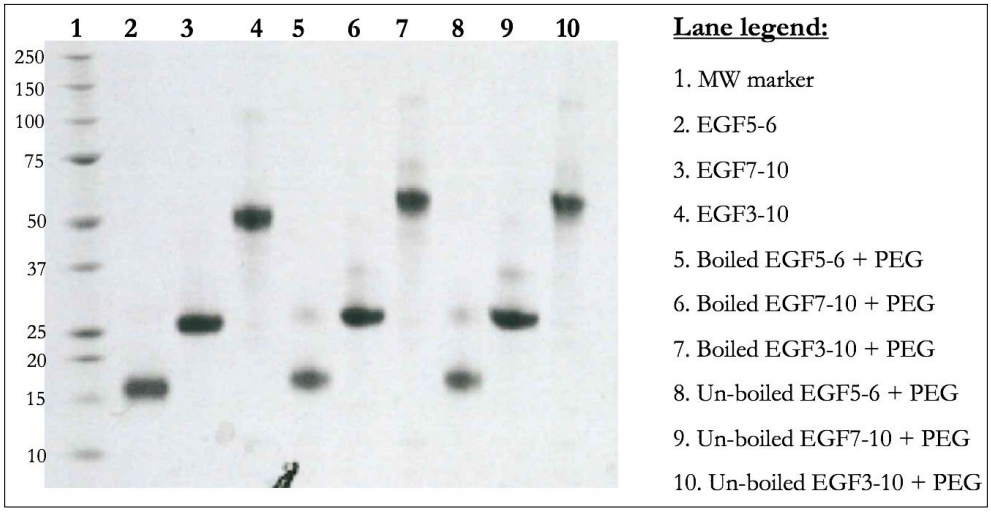


Figure 4.24. PEGylation of Refolded PfRipr Proteins: 3 μg of each protein was analysed on SDS-PAGE using 1 x RSB for each sample. Control samples are in lanes 2 – 4, denatured samples are in lanes 5 – 7, and non-denatured samples are in lanes 8 – 10.

4.4.2.5 Nuclear Magnetic Resonance Spectroscopy

NMR is a powerful technique that can be utilized for the assessment of folding of protein molecules (Kwan, Mobli, Gooley, King, & Mackay, 2011). A common experiment being the 1D ^{1}H-NMR spectrum that contains signals for each non-degenerate hydrogen atom. A folded protein will possess a well-dispersed set of signal peaks as opposed to broad signals indicative of an unfolded/denatured protein. Additionally, evidence of secondary structure can be elucidated, such as the presence of β-sheet. For these reasons initial NMR analysis was carried out on

the refolded EGFs 5-6, 3-6 and 7-10; EGF3-10, at over 50 kDa, was deemed too large for NMR-based analysis.

An initial 1D NMR of EGF5-6 (figure 4.25) reveals evidence of folding due to the good peak dispersion seen, particularly in the aromatic, side chain and backbone amide regions, as well as the carbon-associated hydrogens of the side-chains. Close examination of the amide-proton region at 8.2 - 9.8 ppm revealed ^{1}H peaks with good dispersion; this suggests that the protons are in unique chemical environments due to a well folded tertiary structure of the protein. The peaks are generally sharp, consistent with a lack of self-association or intermediate exchange between conformers. Note, the protein does not contain any Trp residues.

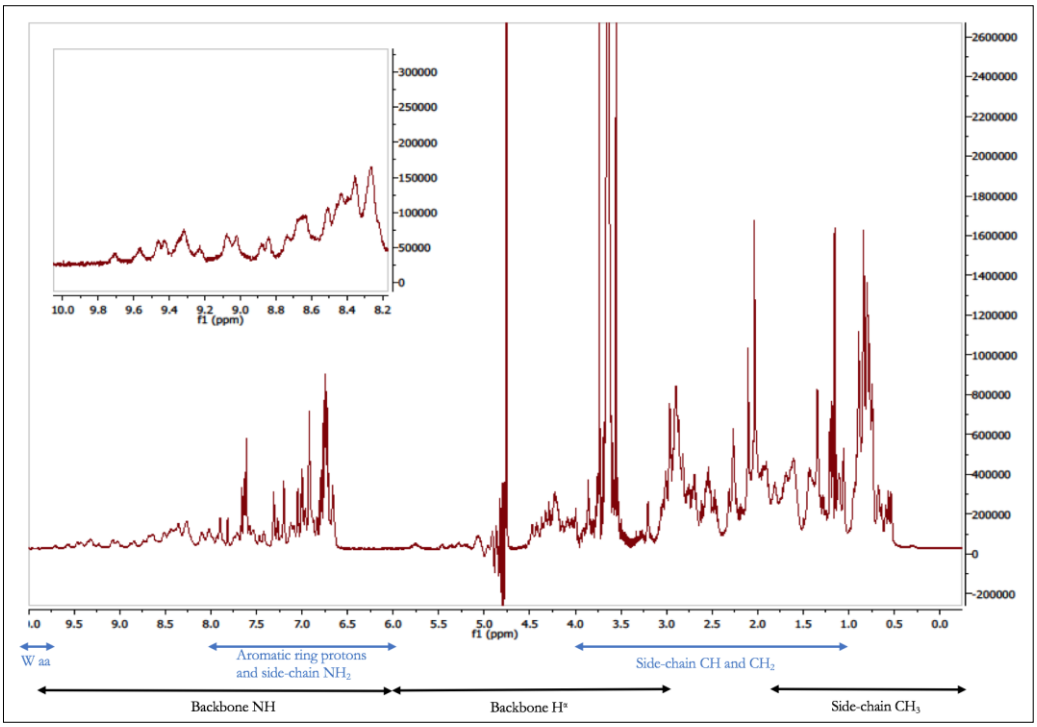


Figure 4.25. 1D ^{1}H-NMR of EGF5-6 in Tris/NaCl: 150 µM sample of EGF5-6 in 20 mM Tris pH 8 + 200 mM NaCl and 10% D_2O. 64 scans taken at an experimental temperature of 25 °C. Inlaid in the figure is a zoom-in of the amide proton region displaying highly dispersed peaks. Abbreviations: W- Tryptophan, aa- amino acid. Figure developed by Roma Galloway and the author.

In order to increase the signal-to-noise ratio and to lessen the effect of the high NaCl concentration (that degrades sensitivity) and artefacts generated from the water and Tris (which was potentially obscuring further peak analysis around 4.7 - 5 ppm), a second sample was produced and buffer-exchanged via dialysis into 1 × PBS from the original Tris/NaCl buffer. A second sample was produced rather than buffer exchanging the existing sample due to the presence of dimer formation over time (data not shown). A sample of around twice the concentration was used with lowered pH from 8 to 7.5 (to sharpen the amide-proton signals by reducing their exchange rate with solvent water). The NMR spectrum of this sample (figure 4.26) shows some potential contamination of Tris (indicated by the peak around 4.7 ppm, likely from around 2 mM of Tris remaining) indicating insufficient buffer exchange, however this resulted in no significant issues with interpretation of the rest of the spectra. Some α-H signals can be seen between 4.8 and 5.5 ppm, which are indicative of the presence of α-protons in a β-sheet in the protein. Narrower peaks were also generated by the slight lowering of the pH from 8 to 7.5 due to a reduced proton exchange in the backbone amide region, particularly around 8.5 ppm.

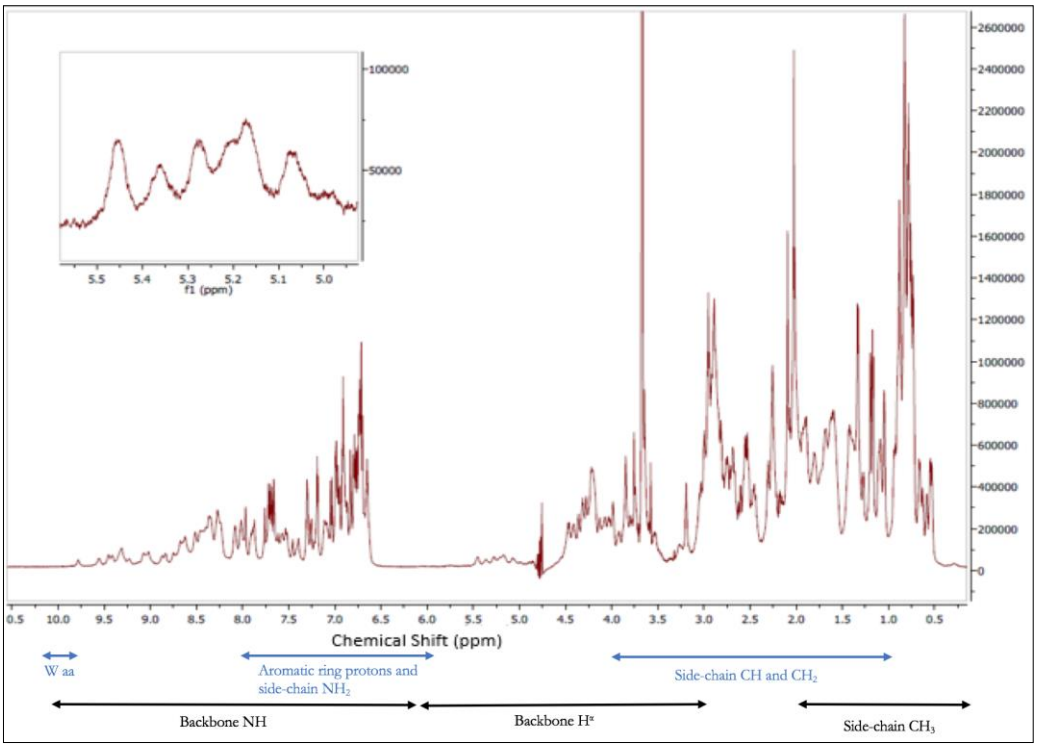


Figure 4.26. 1D ^{1}H-NMR of EGF5-6 in 1 × PBS: 280 μM EGF5-6 in 1 × PBS pH 7.5 (with 40 mM NaCl) + 10% D_2O (and ~2 mM Tris contaminant). Inlaid in the figure is a zoom-in of the 5 – 5.5 ppm region of the spectra. The experiment acquired 64 scans at 25 ºC. Figure adapted from (Galloway, 2016).

Next, the effect of temperature on the 1D NMR spectra of EGF5-6 in 1 × PBS was investigated. A sample was dialyzed into an acidic pH (pH 6.5) PBS buffer to further improve spectra quality due to lower proton exchange in the amide region further. However, an initial sample that was prepared in 1 × PBS without NaCl (to evaluate the effect of removing salt on the sample) ended up precipitating out of solution when analysed at 37 ºC. Due to this it was decided to keep the salt concentration at 80 mM. The results in figure 4.27 show that as the temperature increased from 25 ºC, through 32 ºC to 37 ºC, further peak separation is revealed. This extra resolution is particularly evident around the region of 9.2 – 9.7 ppm, in addition to the peak/s at ~8.4 ppm. The increased sharpness of the spectra is presumably the result of increased molecular tumbling of the protein.

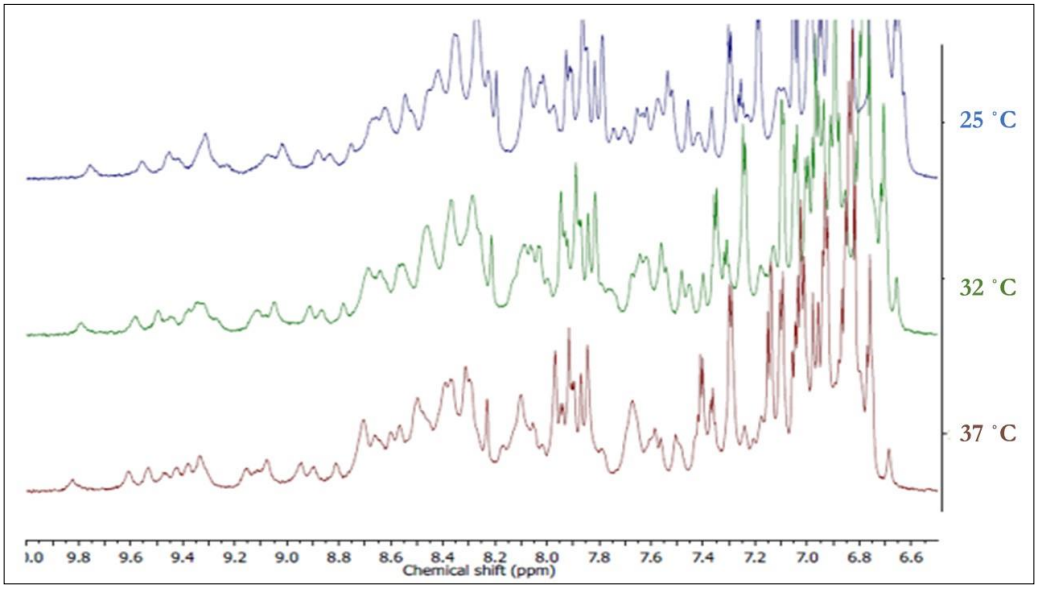


Figure 4.27. 1D ^{1}H-NMR of EGF5-6 Across Varying Temperatures: The 230 μM sample of EGF5-6 (buffer: 1 × PBS pH 6.5, 80 mM NaCl, 2 mM Tris and 10% D_2O) analysed at 25 ºC (purple), 32 ºC (green), and 37 ºC (maroon).

The proton NMR spectrum of a sample of refolded EGF3-6 is shown in figure 4.28. Additional very strong and sharp peaks such as those seen at ~8.5 ppm and 0 ppm were likely due to the presence of protease inhibitors added to the sample for when it was kept in storage at 4 °C. As in the spectrum of EGF5-6, good peak dispersion was seen, including in the backbone amide region. Evidence of β-sheet in the protein was likewise seen at around 5 ppm. It seems likely that most of the EGF3-6 sample was folded but the quality of this spectrum is not as high as that of EGF5-6. Before more NMR-based analysis of EGF3-6 was able to be carried out, the sample was used for other assays. Instead of producing more, the choice to move to a *Pichia*-based expression system was made.

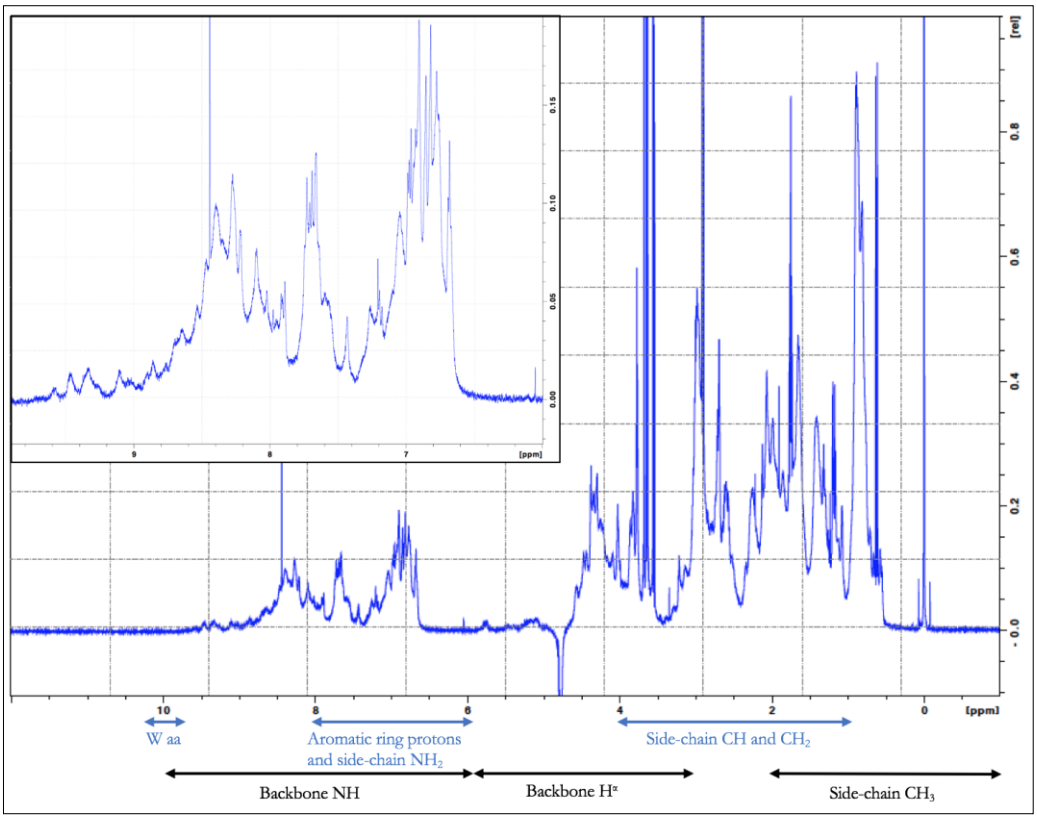


Figure 4.28. 1D ^{1}H-NMR of EGF3-6 in 1 × PBS: 171.5 µM sample of EGF3-6 (in 1 × PBS, pH 8 + 10% D_2O + Roche protease inhibitor cocktail) produced in the pET45b+ plasmid was analysed on an 800 MHz NMR spectrometer. Inlaid in the figure is the zoom-in of the amide region. Experiment carried out at 25 ºC with 64 scans taken. As expected, no tryptophan peak was seen as the protein does not contain any W residues.

Next, 1D-NMR analysis was performed on the refolded EGF7-10 sample (figure 4.29). As was the case with EGF3-6, additional strong peaks at ~8.2 ppm and 0 ppm were likely due to the presence of protease inhibitors added to the sample for when it was kept in storage at 4 °C. Some contaminating glycerol is present at 3.8 ppm, likely from the concentrator tubes used in protein sample preparation. The presence of well-dispersed α-H signals in the backbone region indicated the presence of β-sheet as expected. This, along with the overall sharp, clear dispersion of peaks in both the carbon and amide region demonstrates that most of the protein is likely to be folded.

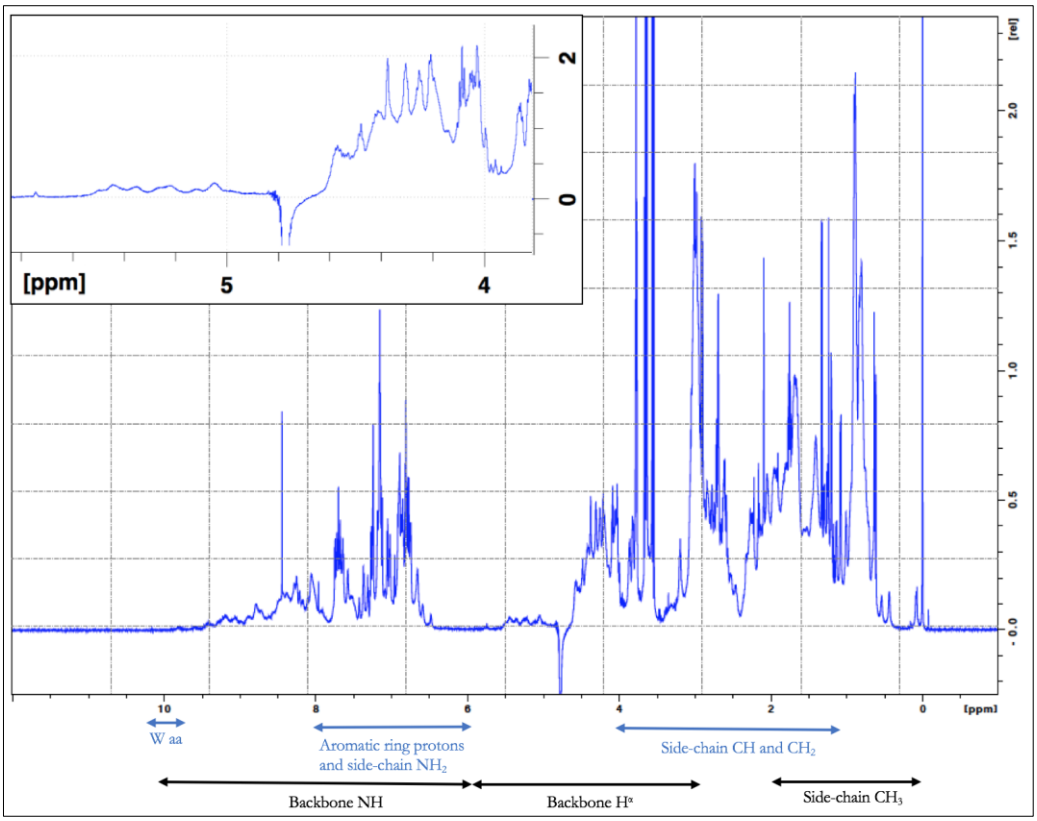


Figure 4.29. 1D ^{1}H-NMR of EGF7-10 in 1 × PBS: 129 μM (in 1 × PBS, pH 8 + 10% D_2O + Roche protease inhibitor cocktail) produced in the pET45b+ plasmid was analysed on an 800 MHz NMR spectrometer. Inlaid in the figure is the zoom-in of the backbone H^{α} region.

Next, the effect of both temperature and pH were investigated on a new sample of refolded EGF7-10 that had been produced. The best spectrum was obtained at 310 K, with mild improvement on peak sharpness, particularly at around 4.25 ppm (figure 4.30). A reduction of pH had a more significant effect on resolution, with the sample in pH 4.5 showing better signal dispersion in the backbone H^{α}-region (figure 4.31).

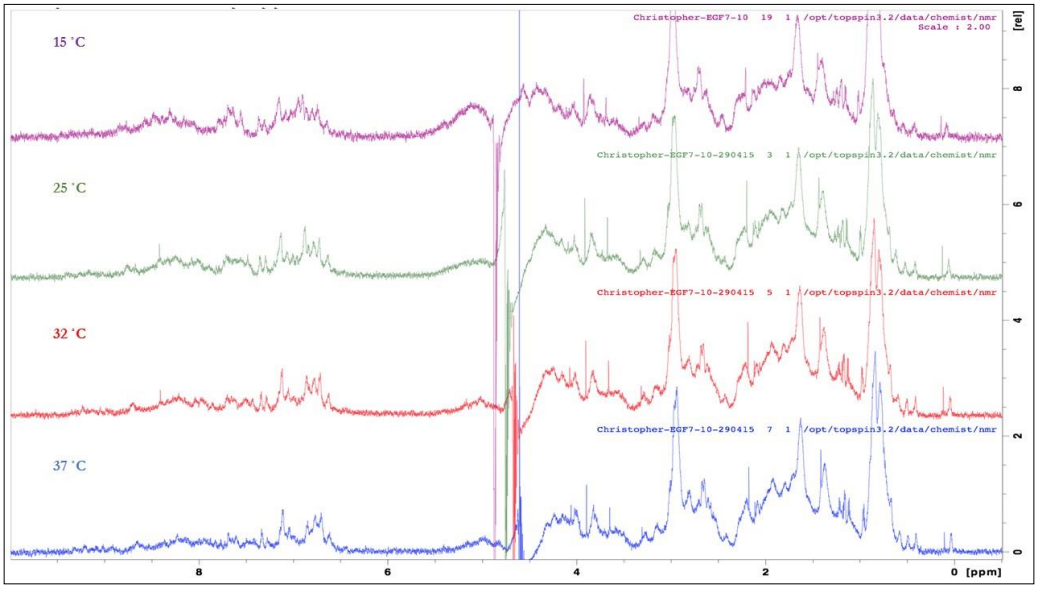


Figure 4.30. 1D ^{1}H-NMR of EGF7-10 Across Varying Temperatures: The 123.7 µM sample of EGF7-10 (buffer: 1 × PBS pH 7.5, 80 mM NaCl, and 10% D_2O) analysed at 15 ºC (violet), 25 ºC (green), 32 ºC (red), and 37 ºC (blue) for 32 scans.

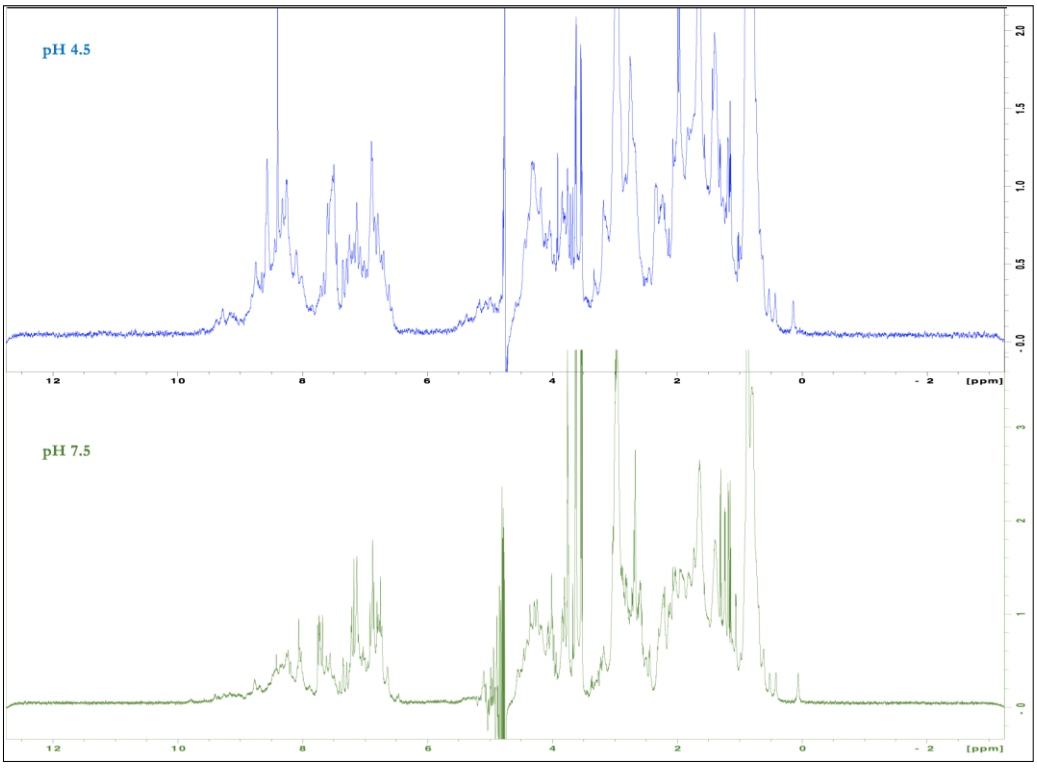


Figure 4.31. 1D ^{1}H-NMR of EGF7-10 pH Screen: The 123.7 µM EGF7-10 sample at pH7.5 was in 1 × PBS pH 7.5, 80 mM NaCl, and 10% D_2O (green spectra). The 111 µM EGF7-10 sample at pH4.5 was in 1 × PBS (pH 4.5 via adjustment with acetic acid), 80 mM NaCl, and 10% D_2O (blue spectra). Experiment carried out at 25 ºC for 64 scans.

2D NMR experiments on the non-isotopically labelled EGF5-6 refolded protein was decided upon to provide further evidence for protein folding, but to inform on prospective future experiments with isotopically-labelled sample for further structural studies. A natural abundance ^{15}N-^{1}H heteronuclear single quantum coherence (HSQC) NMR spectrum was obtained using the EGF5-6 sample prepared in PBS. The results (figure 4.32) gave relatively limited information, however the dearth of overlapping peaks indicated that the sample had not aggregated over the ~40-hour period of the experiment. With respect to evidence for folding, there are well dispersed peaks as would be expected that also corroborate the evidence for folding from the 1D experiments. Increasing the temperature further increased the sharpness of the signals, though the sample was analysed for less time so fewer signals were observed.

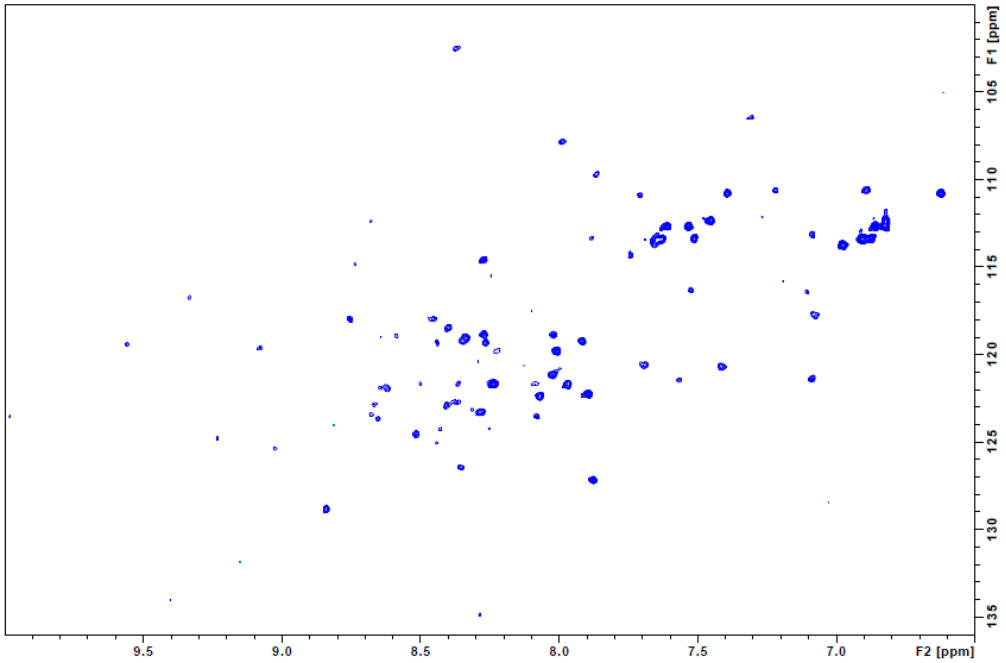


Figure 4.32. Natural Abundance ^{15}N-^{1}H HSQC NMR Spectrum of EGF5-6: Signals were from a 280 µM sample (in 1 × PBS pH 7.5, 40 mM NaCl, 2 mM Tris and 10% D_2O) analysed at 25 ºC over 40-hours.

A ^{1}H-^{1}H nuclear Overhauser effect spectroscopy (NOESY) experiment was performed to reveal through-space proximities between protons as another way to assess folding (figure 4.33). A well-populated NOESY spectrum was obtained for the EGF5-6 sample with good dispersion. Cross-peaks between, for example, α-protons and NH protons and between aromatic and methyl protons are definitive evidence for a folded proton. The lack of clumps of cross-peaks in this spectrum suggests there is little unfolded material in the sample. Overall, the results for EGF5-6 were very promising and suggested that it would be worthwhile to generate a ^{15}N-labelled sample suitable for more sophisticated NMR analysis.

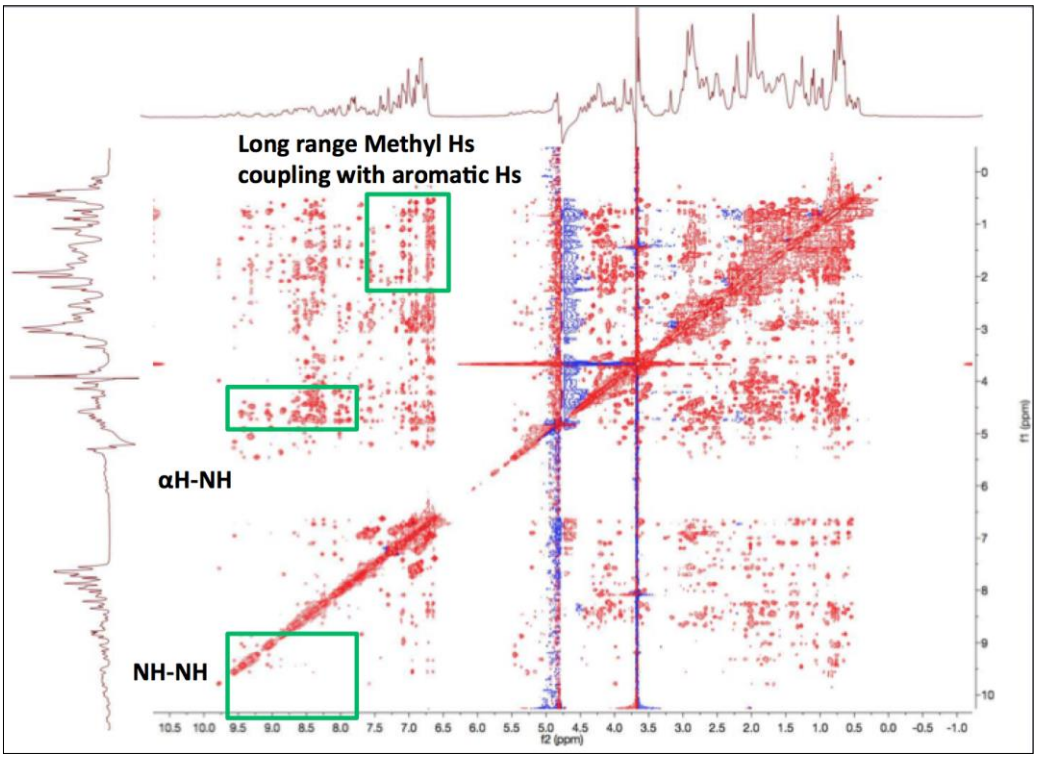


Figure 4.33. ^{1}H-^{1}H NOESY NMR Spectrum of EGF5-6: The 280 μM sample (in 1 × PBS pH 7.5, 2 mM Tris, 40 mM NaCl, and 10% D_2O) was exposed to a temperature of 25 ºC during the experiment. Key regions of proton interactions are highlighted in the figure. Figure adapted from (Galloway, 2016).

HSQC and NOESY experiments were initially carried out on an unlabelled sample of EGF7-10, however the quality of the spectra was poor (data not shown for HSQC experiments; see figure 4.34 for the NOESY). This may be due to the protein being having an elongated rod-like shape (i.e. a large Stokes radius) leading to slow molecular tumbling. Moreover, low sample concentration resulting from sample loss during preparative steps and proton exchange with the H_2O in the buffer. While some cross-peaks in the NOESY spectrum suggested protein folding (such as those in the long-range methyl region), they were relatively weak, especially contrasted to equivalent peaks seen for the EGF5-6 spectrum in figure 4.33. To reduce the size of EGF7-10, attempts were made to cleave the polyhistidine tag from the N-terminus of the protein via EK (which comprised 2.2 kDa of the full size of the recombinant protein, or roughly 9.1% of the total amino

acid sequence). Despite varying enzyme concentration, time of EK exposure and temperature of the reaction, this was ultimately unsuccessful (data not shown). Instead of re-cloning the EGF7-10 gene into a pProExHTA expression system (containing a TEV cleavage site for removal of the polyhistidine tag), it was ultimately decided to move to a *Pichia*-based expression system where refolding would not be necessary.

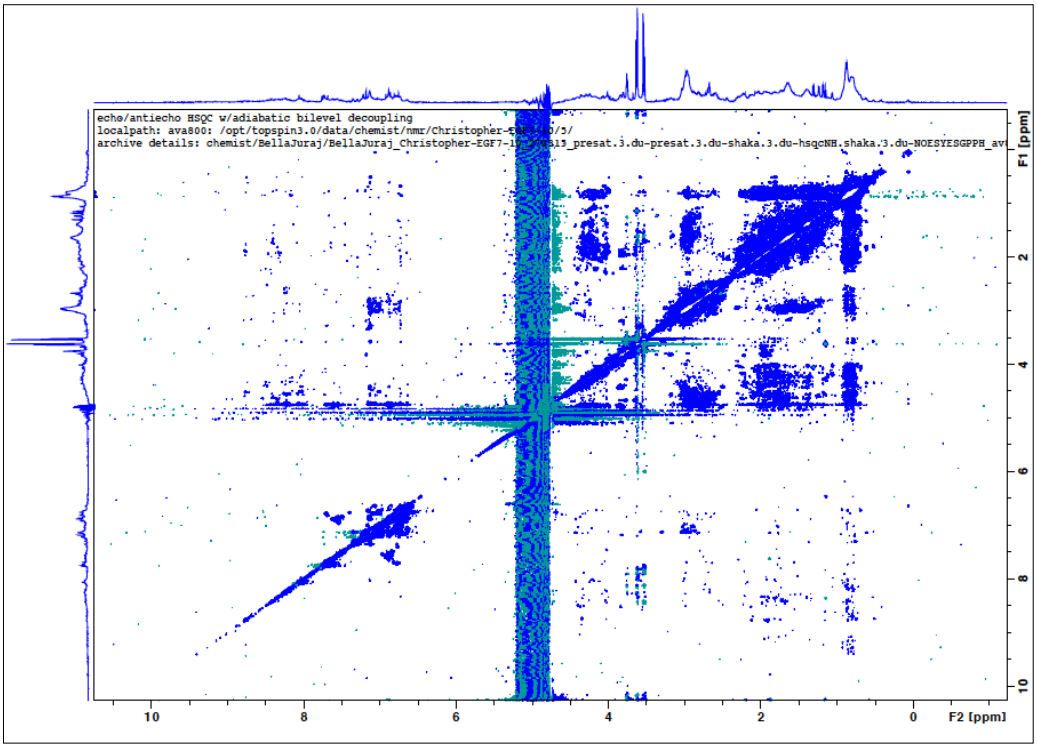


Figure 4.34. ^{1}H-^{1}H NOESY NMR Spectrum of EGF7-10: The 123.7 μM EGF7-10 sample at pH7.5 was in 1 × PBS pH 7.5, 80 mM NaCl, and 10% D_2O was exposed to a temperature of 25 °C during the experiment. Poor signal intensity is revealed for long-range methyl-associated protons, aromatic proton interactions and the interaction between the backbone amide protons and α-H through space.

It was attempted to produce EGF5-6 from *E. coli* grown in minimal media *en route* to preparing an isotopically labeled sample. This was unsuccessful likely due to the glycerol stock being unreliable. Rather than perform a fresh transformation of

BL21 (DE3) *E. coli*, it was decided to move exclusively into a *Pichia* system to produce proteins containing the 5th and 6th EGF-like domain of PfRipr.

4.5 Binding Studies between PfRipr Constructs and PfRh5

4.5.1 Methods for PfRipr – PfRh5 Binding Studies

4.5.1.1 ELISA Assays

Flat-bottomed 96-well plates (Maxisorp, Nunc) were coated with a 1:2 serial dilution starting with 200 µl of 64 µg/ml recombinant PfRipr protein made up in PBS (across eight wells and set up in triplicate). Controls in triplicate were also set up including a PBS-only negative control, PBS-only for exposure to antibodies only (negative control), protein-only positive controls, and a 1:2 dilution series of 100 µl BSG in PBS at 8 mg/ml (positive control). Plates were incubated overnight at 4 °C.

Following incubation, plates were washed 3 x with PBS-T, blocked with 200 µl of 5% w/v dry skimmed milk powder in PBS-T for at least one-hour at room temperature. After another wash step, 100 µl of 5 µg/ml PfRh5 (prepared in 1 × PBS) was added to all appropriate wells for an incubation period of two-hours at room temperature. Plates were then washed and 50 µl of primary antibody added (a 1:10 000 dilution of mouse anti-PfRh5 antibody made up in blocking solution; for the positive control only samples, rabbit anti-PfRipr, rabbit anti-BSG, or mouse anti-pentahistidine antibodies were used at a 1:2000 dilution).

After a 90-minute incubation and subsequent washing, 50 µl of the secondary antibody was added (either anti-mouse HRP or anti-rabbit HRP at a 1:2000 dilution). Plates were further incubated for one-hour and washed thrice with PBS-

T then twice with HTPBS. 100 µl of equal parts TMC peroxidase substrate and peroxidase substrate solution B (Sigma Aldrich) were added to each well for a one- to ten-minute development in the dark. The reaction was quenched with 50 µl of phosphoric acid (from stock: 17 ml of pure H_3PO_4 in 250 ml pure H_2O). Plates were then analysed at 450 nm absorbance on a PerkinElmer Fusion Alpha plate reader.

4.5.1.2 Coimmunoprecipitation Assays

For the co-immunoprecipitation (coIP) assay between PfRipr proteins and PfRh5, two methods were employed. The first method involved the use of 10 µl of anti-Flag beads (donated by Dr. Lin Chen at the Walter and Eliza Hall Institute) per protein sample to be used. 10 µg of each protein was added to 35 µl of PBS, mixed, then an aliquot of 5 µl of this was kept to run on a gel as the input sample. The protein samples were added to anti-Flag beads and mixed overnight at 4 °C. After incubation, samples were centrifuged at 500 × g for 30-seconds at 4 °C, collecting the unbound fraction to run on a gel. Three washes with 1 ml of PBS were carried out, followed by two elution steps using 20 µl of elution buffer (containing Flag peptide), further centrifugation, followed by loading 20 µl of the elution fractions onto a gel.

The second method consisted of the following steps: At 4 °C, 1 µg of each protein was mixed together and incubated overnight. Control samples were to receive no primary antibody in order to evaluate any non-specific association of the protein with protein G beads. Rabbit polyclonal anti-PfRh5 antibody in a 3:1 ratio was added and incubated for two-hours at room temperature. An aliquot of 25 µl of protein G beads from Qiagen (after washing them in PBS) were added and mixed at for two-hours at 4 °C. Samples were centrifuged at 1500 × g for one-minute, then spun at 10 000 × g for ten-seconds, collecting unbound fraction. 3 - 5 x 200

µl PBS washes were carried out (keeping the wash fractions). Next, 25 µl of 1 x RSB was added, samples were boiled for two-minutes, then centrifuged at 10 000 × g for one-minute. An SDS-PAGE and a Western blot were carried out as per the general methods outlined in section 2.2.1, using 100 ng of PfRh5 and PfRipr as standards, probing with mouse anti-pentahistidine antibodies. Where magnetic beads were used, they were Dynabeads (ThermoFisher) used with a magnetic stand (FastGene MagnaStand 1.5) donated by Dr. Anthony Hodder at the Walter and Eliza Hall Institute.

4.5.1.3 Gel Filtration Binding Assays

A sample consisting of 50 - 100 µg of each protein was analysed both individually and mixed together (for either two-hours at room temperature or overnight in a cold room) then loaded onto on a calibrated Superdex 200 10/300 GL SEC column with the assistance of Dr. Lin Chen. The contents of each collected fraction were analysed by SDS-PAGE and Western blot using the methods outlined in section 2.8.

4.5.2 Results for PfRipr – PfRh5 Binding Studies

4.5.2.1 ELISA Assays

Several attempts to optimise the ELISA assay for protein-protein binding were made using a range of protein concentrations, wash conditions, antibody dilutions and incubation times (data from optimization phases not shown). Eventually, promising results were obtained by using the method described in section 3.5.1.1 (see figure 4.35). It should be noted that EGF3-10 was not used in the assay as it had not been developed successfully at the time these experiments were performed. None of the EGF-rich PfRipr proteins tested gives a dose response

curve (like that seen with the BSG then Rh5 positive control), except EGFs 5-6 and 3-6, appearing to bind with a similar response to each other. The binding here appears to be around one-fifth of the intensity seen in the positive control. However, the control PfRh5 samples that were exposed to PBS instead of PfRipr then had regular antibody detection gave an average signal of 0.378 +/- a standard deviation of 0.229; potentially quite high but with a wide deviation. Due to this high background and its potential implications for the results shown, it was decided to repeat the assay.

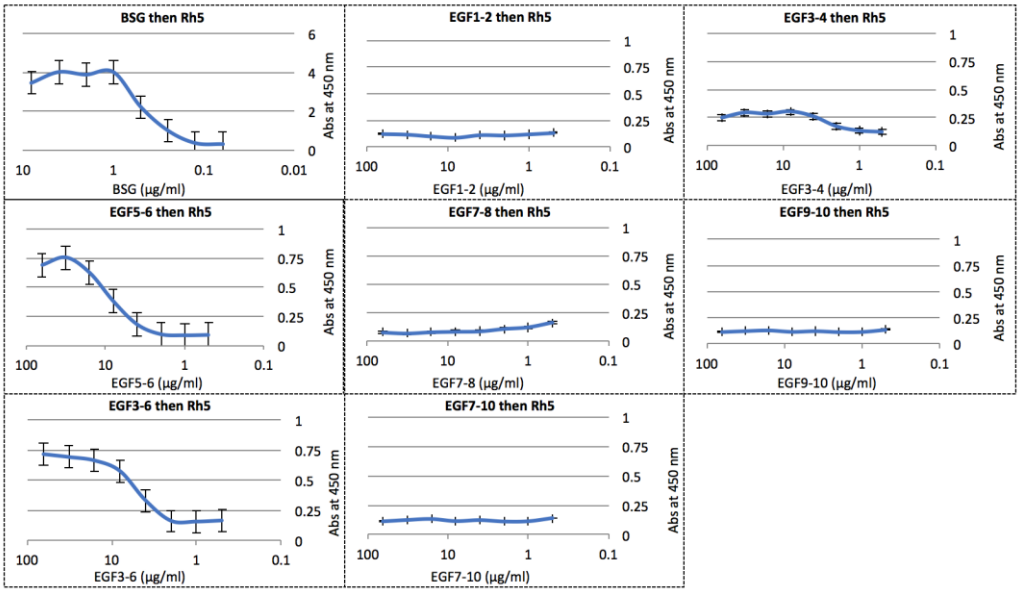


Figure 4.35. ELISA Binding Assay of PfRh5 and PfRipr EGF-Rich Proteins: A reference positive binding result is shown in the "BSG then Rh5" panel, in which a concentration series of BSG was exposed to 100 µl of 5 µg/ml PfRh5. The error bars represent the standard error. All results have been adjusted for the background noise of the negative controls. Positive controls for each protein gave strong absorbance readings (~1.5) at 450 nm, whilst the anti-PfRh5 antibodies did not appear to cross-react with PfRipr control samples as the signal remained in the background noise. However, these antibodies gave rise to an absorbance signal of 0.518 within the BSG control well likely due to using too much protein at 8 µg/ml rather than 1 µg/ml (data not shown). All EGF-rich PfRipr proteins used in the assay were from the soluble material produced in section 3.2, rather than refolded material.

Unfortunately, this high background issue persisted despite attempts to optimize the assay. These included the use of a variety of blocking agents (BSA, SuperBlock™ blocking reagent and gelatin mix), temperatures (room temperature, cold room, 30 and 37 °C), and blocking time (6-, 24- and 48-hours). Except for EGF3-4 (which perpetually aggregated out of solution), all proteins appeared intact when analysed via SDS-PAGE under R and NR conditions, and via mass spectrometry (data not shown); whenever any proteins appeared degraded, a fresh batch was prepared. Once refolded proteins had been produced, they were used in numerous repeats of the assay, albeit still with negative results (i.e. a lack of binding that was originally seen) aside from the positive control of PfRh5 binding BSG. This included attempts whereby the plate was coated with PfRh5 and then PfRipr EGF-rich proteins were titrated in with detection using rabbit polyclonal anti-PfRipr antibodies (this resulted in a high background). Nonetheless, the possibility remained that PfRipr constructs (EGF 3-6 and EGF 5-6), i.e. those containing the 5th and 6th EGF-like domains, bound to PfRh5. To assess their possibility further, alternative binding experiments were attempted.

4.5.2.2 Coimmunoprecipitation Assays

The initial coIP assay for the PfRh5:EGF5-6 interaction utilized controls of Flag-tagged PfRh5, refolded EGF5-6 and a mix of PfRh5 and EGF5-6 with anti-Flag beads (figure 4.36). Here, each control protein is shown at their correct relative sizes in each input sample. The positive control Flag-tagged PfRh5 protein however, failed to bind to the anti-Flag bead resin (rendering the experiment itself useless), with what appears to be the entirety of the protein found in the unbound fraction and none in the eluted fraction. This was postulated to be potentially due to proteolytic degradation of PfRh5 (some faint bands below the PfRh5 input sample can be seen) which could result in loss of the Flag-tag.

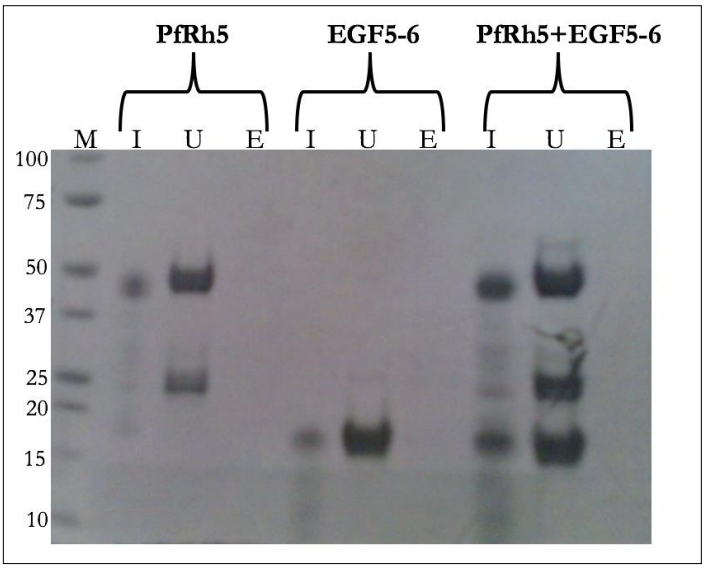


Figure 4.36. Initial CoIP Test Between PfRh5 and EGF5-6: 10 µg of each protein (3.3 µl of Flag-tagged PfRh5 and 7.57 µl of EGF5-6) in 35 µl of 1 × PBS (pH 8) was used for the assay, with 5 µl being used for the input (I) samples on the SDS-PAGE. The sample binding was carried out over three-hours at room temperature on a rotary mixer. As expected, the non-Flag-tagged EGF5-6 was only found in the unbound fraction (U). Other abbreviations: E- eluted fraction.

In an attempt to rectify this, a pure sample of PfRh5 that had previously been verified to contain the Flag-tag (via Western blot) was donated by Dr. Lin Chen. The coIP experiment (figure 4.37) was carried out with overnight incubations in an attempt to ensure a sufficient amount of time was allowed for the binding of PfRh5 to the resin. In this experiment, PfRh5 was not found in the unbound fraction but nor was it seen in the eluted fraction. This suggested the PfRh5 might have been removed during the wash steps, yet upon analysis of wash fractions by SDS-PAGE, no PfRh5 was detected (results not shown). This may have been due to the large volume of the wash fraction and the consequent limits to detectability of PfRh5. The experiment was therefore repeated but using only one-tenth (0.1 ml) of the previously used volume of PBS for the wash (figure 4.38). Despite an overnight incubation with mixing at 4 °C, much of the Flag-tagged PfRh5 remained unbound, while the majority of the protein was found in the first wash fraction. These results suggest that despite nominally having a Flag-tag, our PfRh5

protein produced did not bind to the resin used. Therefore, this coIP method was unsuitable for further binding studies between PfRh5 and PfRipr proteins.

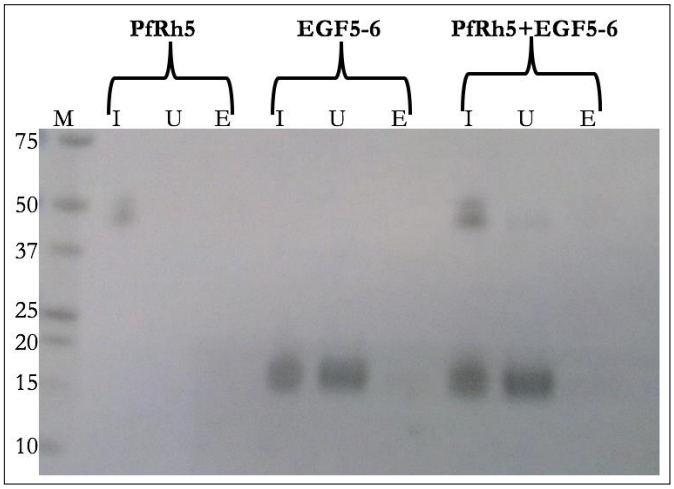


Figure 4.37. Second CoIP Test Between PfRh5 and EGF5-6: Sample binding was carried out overnight at 4 °C. The fractions analysed on the SDS-PAGE gel suggest that the PfRh5 sample has been lost along the way, possibly during the wash steps.

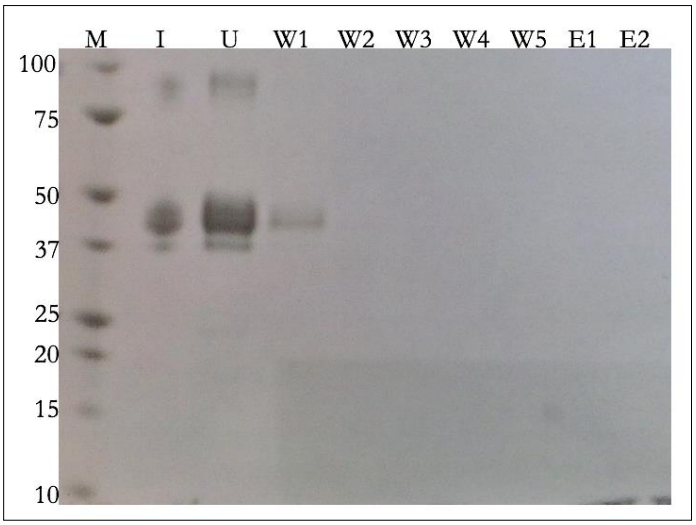


Figure 4.38. CoIP Optimization for PfRh5: This experiment was carried out to troubleshoot and optimize conditions for the anti-Flag beads to capture the Flag-tagged PfRh5 protein (shown with a small amount of dimer present in the input (I) fraction. Each wash step was carried out with 0.1 ml of PBS. The first elution (E1) utilized the Flag peptide, the second (E2) utilized 0.25 M glycine pH 2.6 + 0.75 M Tris pH 9.5.

It was decided to try an alternative coIP strategy, using rabbit polyclonal anti-PfRh5 antibodies (donated by Dr. Lin Chen) to capture the PfRh5, and protein G for antibody capture on the resin. However, whilst assessing the experimental system for any potential flaws, it was revealed that the PfRh5 binds to the protein G beads non-specifically (figure 4.39, lane 3). This is hence an unusable assay with the prospect of false positive results. To try and get around this problem, a method of pre-clearing the PfRh5 sample was attempted, whereby the protein and protein G beads would be co-incubated then centrifuged at high speed, keeping the supernatant with PfRh5 for antibody capture followed by a second incubation with the beads.

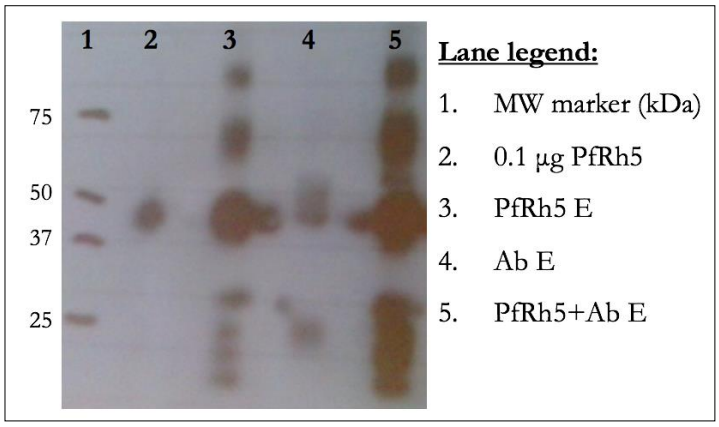


Figure 4.39. Protein G Test for PfRh5 CoIP: This Western blot was developed using a mouse anti-PfRh5 antibody and a goat anti-mouse HRP antibody, both at a dilution of 1:1000. Lane 1 shows marks corresponding to the associated MW markers. Lane 2 is a positive control for PfRh5. Lane 3 corresponds to a sample of PfRh5 incubated with protein G. Lane 4 corresponds to the rabbit polyclonal anti-PfRh5 antibodies incubated with protein G. Lane 5 shows the PfRh5 + the anti-Rh5 antibody. Development time was one-second.

To limit the potential of pipetting error, magnetic Dynabeads were used. The experiment from figure 4.39 was repeated, but this time with pre-clearing that sample as described above. The results in figure 4.40 show that unfortunately the PfRh5 protein still bound to the bead matrix. Due to this remaining a problem with respect to protein-protein interaction studies, it was decided to attempt to

study binding via FRET, BN-PAGE and/or SEC. To note, the positive control for the FRET assay was unable to be established and so this method was abandoned. Further, the BN-PAGE data were inconclusive; therefore, neither of the data for these experiments are sown in this thesis.

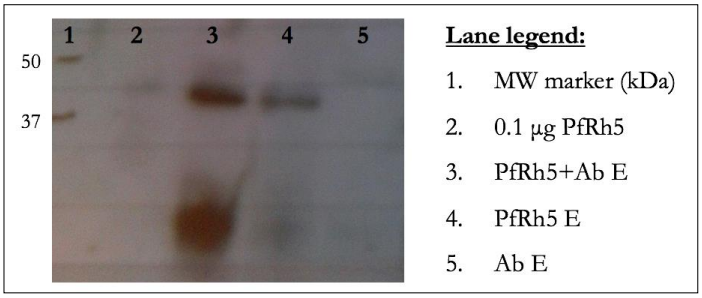


Figure 4.40. Pre-Cleared PfRh5 Test for CoIP: This Western blot was developed using a mouse anti-PfRh5 antibody and a goat anti-mouse HRP antibody, both at a dilution of 1:1000. Development time was 20-seconds.

4.5.2.3 Gel Filtration Binding Assays

SEC was employed to look for evidence of a complex being formed between either EGF3-6 or EGF3-10 and PfRh5. This experiment would be able to at least provide a relatively conclusive answer for if a reasonably strong complex is formed between the EGF-rich region of PfRipr and PfRh5. First, EGF3-6 was loaded onto a calibrated Superdex 200 10/300 GL column followed by PfRh5 to establish the normal peak placement of each protein in the gel filtration elution profiles (figure 4.41). Following a 2.5-hour incubation at room temperature, the PfRh5 + EGF3-6 mix was analysed via SEC (figure 4.42). No significant peak shift was seen that would indicate that the proteins formed a tightly bound complex.

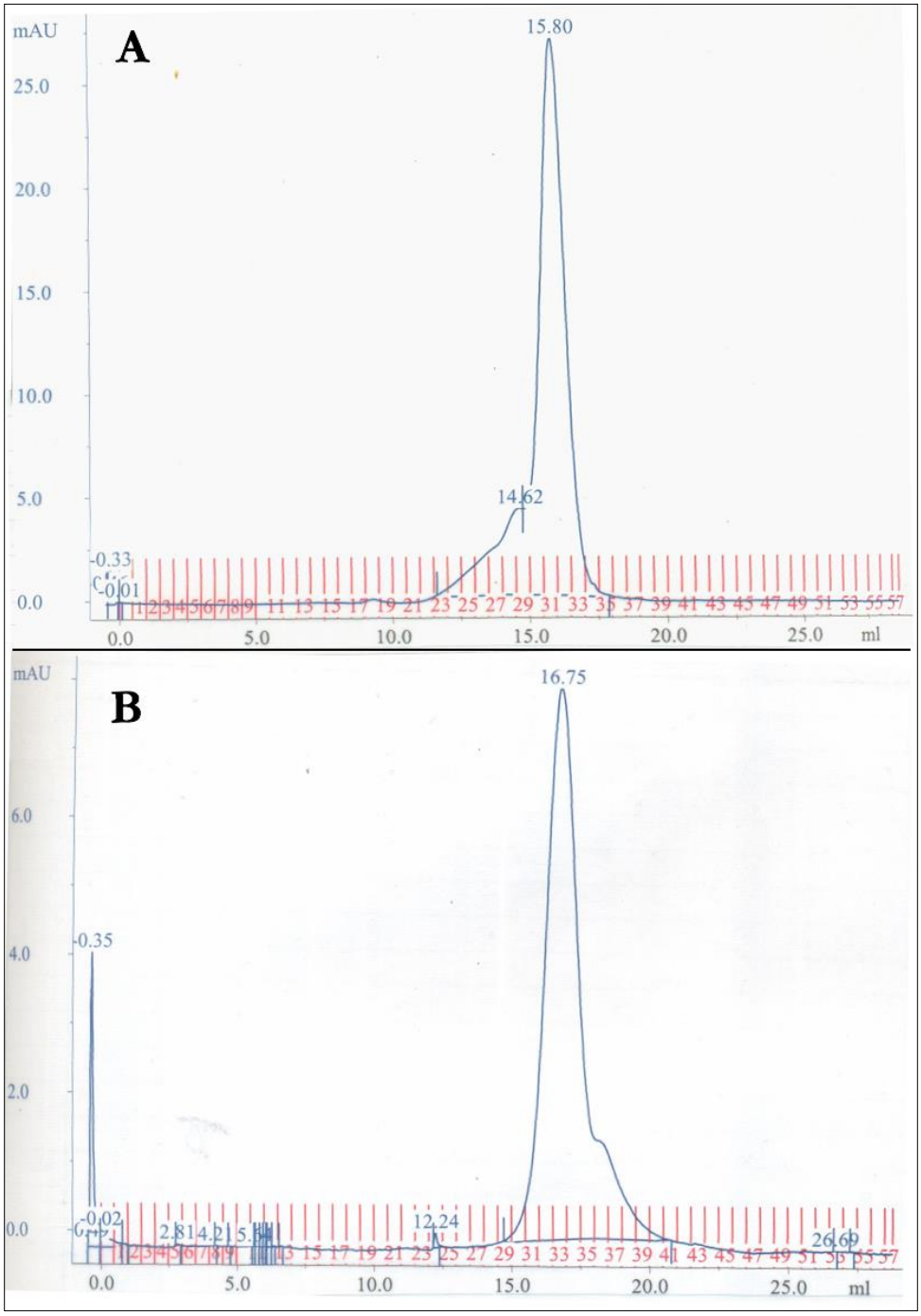


Figure 4.41. SEC of EGF3-6 and PfRh5 on an S200 10/300 GL Column: A. The SEC chromatogram of 100 µg of EGF3-6. A small shoulder peak is seen at 14.62 ml which corresponds to a small amount of dimer, whilst the major peak at 15.80 ml is the monomeric protein. B. The SEC chromatogram of 18.5 µg of PfRh5. The major peak at 16.75 ml is where the monomeric protein elutes, whilst the minor shoulder after this is a small amount of degradation product.

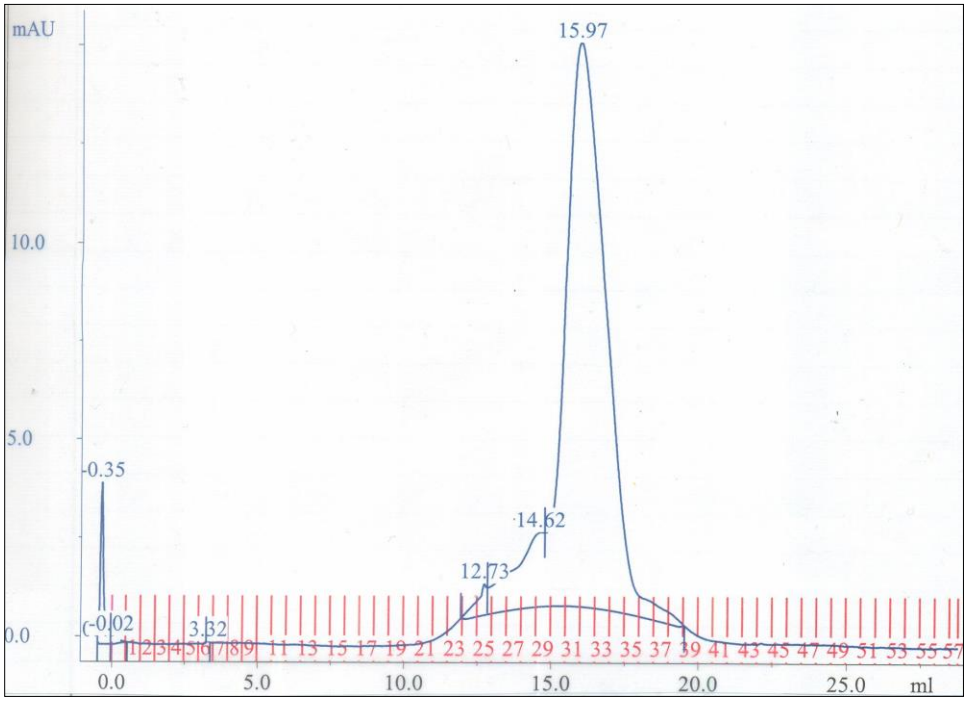


Figure 4.42. SEC of EGF3-6 + PfRh5 on an S200 10/300 GL Column: 50 µg of EGF3-6 was mixed with 18.5 µg of PfRh5 and injected onto the gel filtration column.

To assess if the time of incubation made any difference, a protein mixture was prepared and incubated overnight. A different gel filtration column was used and, each individual protein was analysed as before (figure 4.43). Again however, once the mix of proteins were applied to the gel filtration column, no evidence of a tight complex between EGF3-6 and PfRh5 was seen (figure 4.44)

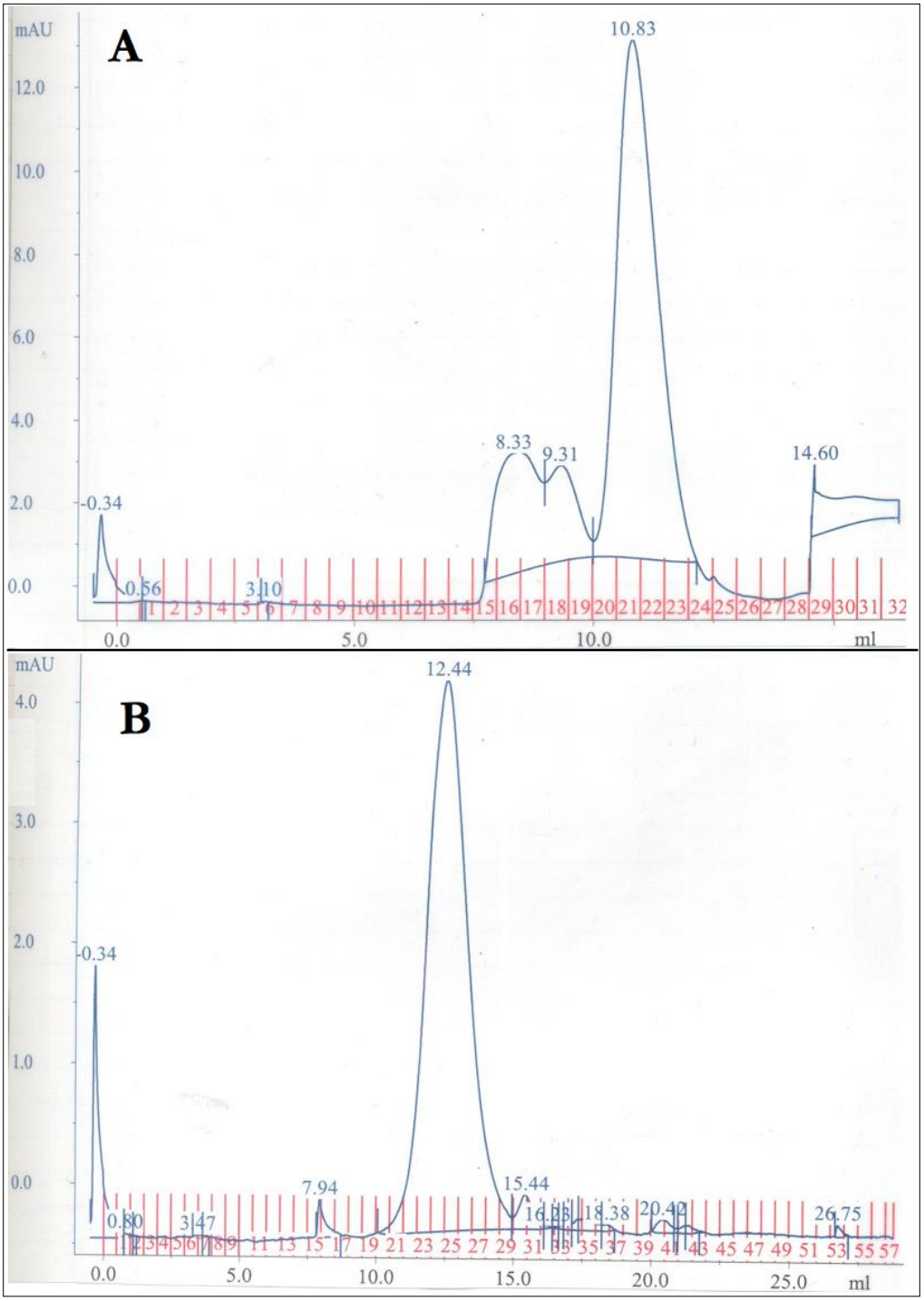


Figure 4.43. SEC of EGF3-6 and PfRh5 on an S75 10/300 GL Column: A. The SEC chromatogram of 100 µg of EGF3-6. Two small shoulder peaks are seen at 9.31 and 8.33 ml which corresponds to small amount of dimer and trimer (respectively), whilst the major peak at 10.83 ml is the monomeric protein. B. The SEC chromatogram of 18.5 µg of PfRh5. The major peak at 12.44 ml is where the monomeric protein eluted.

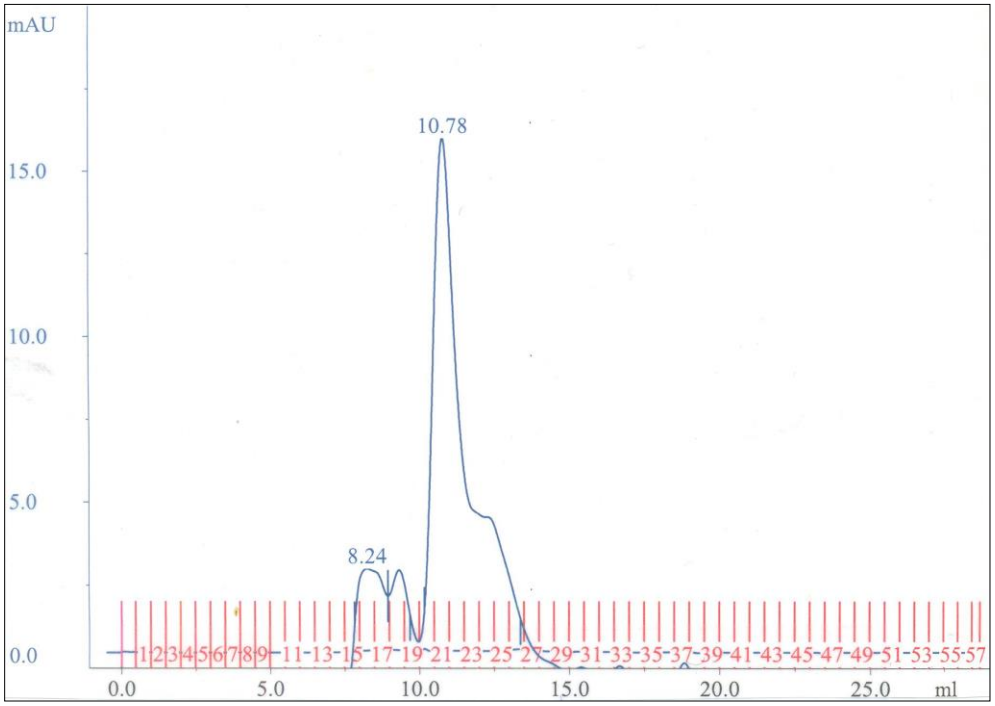


Figure 4.44. SEC of EGF3-6 + PfRh5 on an S200 10/300 GL Column: 50 µg of EGF3-6 was mixed with 18.5 µg of PfRh5 and injected onto the gel filtration column.

A sample of EGF3-10 was used to assess if the entire backend of PfRipr could form a tight complex with PfRh5. If so, this would suggest that the binding sites for PfRh5 might lie outside EGF modules 5 and 6. Each protein was first passed individually over the SEC column (figure 4.45), then they were mixed together, incubated for two-hours at room temperature and passed through the column (figure 4.46).

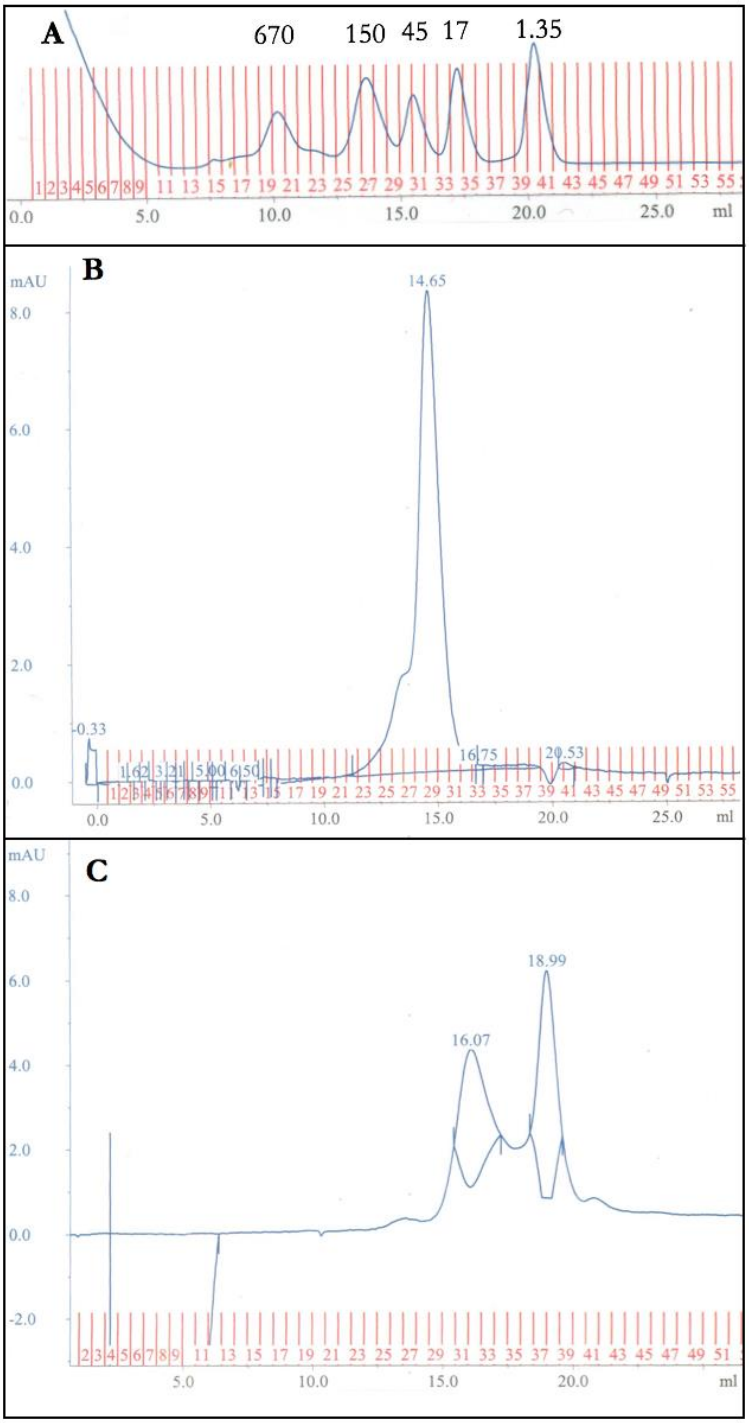


Figure 4.45. SEC of EGF3-10 and PfRh5: A. Molecular weights standards shown (in kDa). B. 16.8 µg of EGF3-10 with the monomeric protein eluting at 14.65 ml as expected. Total sample injection volume was 100 µl. B. 32 µg of PfRh5 with the monomeric fraction eluting at 18.99 ml and dimer around 16.07 ml (both as approximately expected). Total sample injection volume was 100 µl. Column used was an analytical S200 10/300.

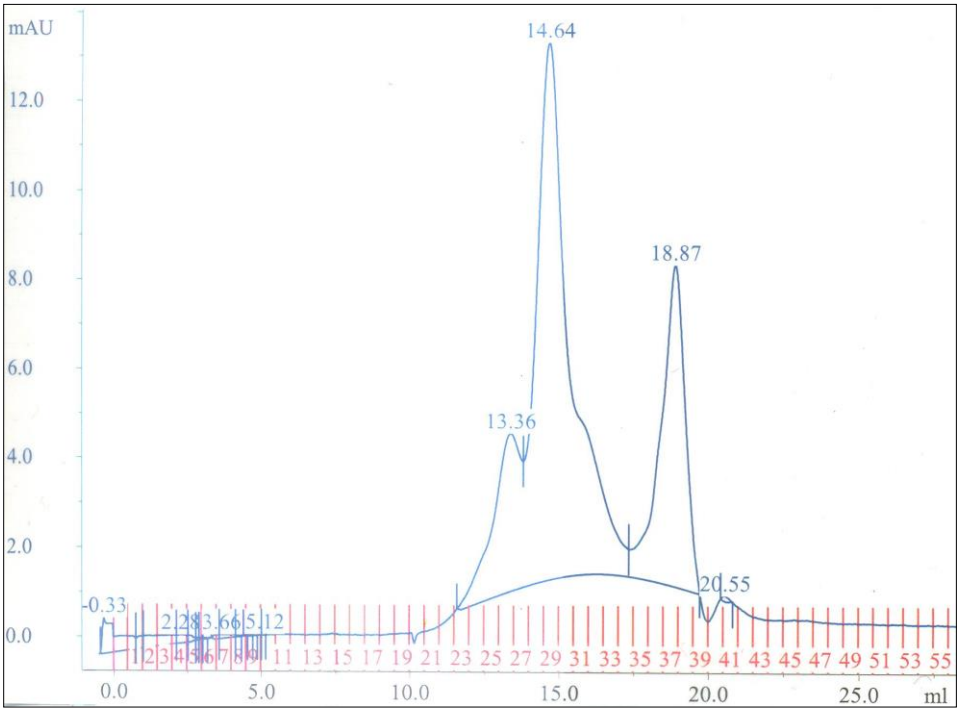


Figure 4.46. SEC of EGF3-10 + PfRh5: 16.8 µg of EGF3-6 was mixed with 32 µg of PfRh5 and injected onto the gel filtration column. Total sample injection volume was 200 µl. Column used was an analytical S200 10/300.

Based on the chromatogram in figure 4.46, no significant evidence was obtained in support of a tight complex between EGF3-10 and PfRh5. However, the twofold increase in mAU of the shoulder peak at around 13 ml (corresponding to a molecular weight of approximately 150 kDa) and its increased prominence compared to the shoulder for EGF3-10 in figure 4.45 B, motivated an analysis of the eluted fractions by SDS-PAGE and Western blot (figure 4.47). It was hoped that fractions 25 and 26 would contain both partners; however, no bands could be detected. Whilst there is a band around 100 kDa for the anti-PfRipr and anti-pentahistidine detected samples in fraction 26, the same is not the case when detecting with anti-PfRh5 antibodies. These results therefore do not support formation of a tight complex between EGF3-10 and PfRh5.

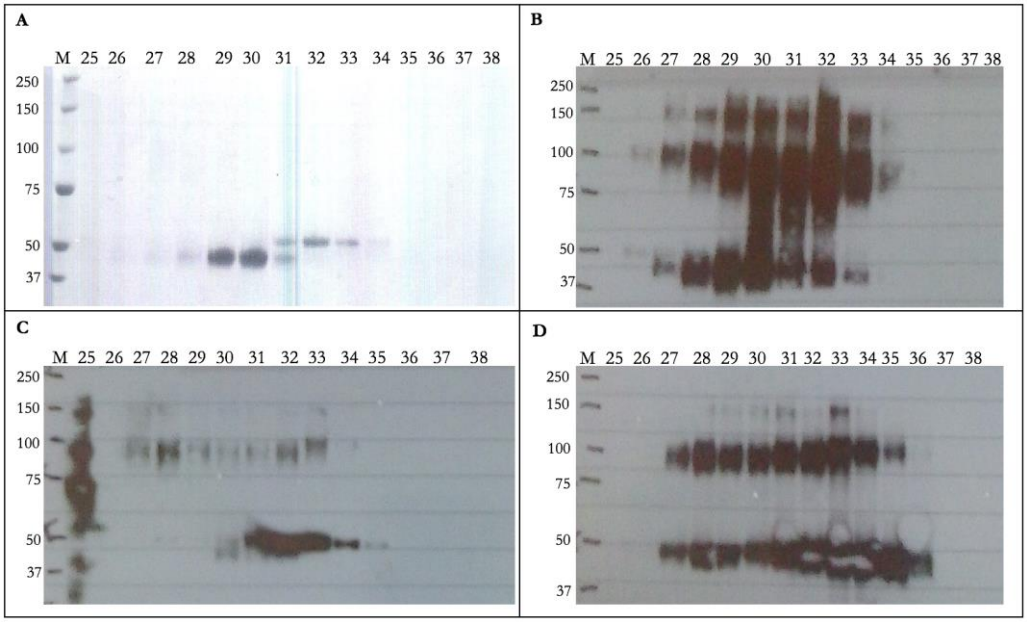


Figure 4.47. SDS-PAGE and Western Blot Analysis of EGF3-10 + PfRh5 SEC Fractions: A. Coomassie-stained gel of fractions 25 – 38 analysed with 1 x NRSB. B. Anti-PfRipr antibodies used for detection. C. Anti-pentahistidine antibodies used for detection. D. Anti-PfRh5 antibodies used for detection. Western blots in B – D all developed within two-seconds.

4.6 PfRh5 and BSG Antibody Screens

4.6.2 Methods for PfRh5 and BSG Antibody Screens

For initial antibody screening, 100 µl of 0.5 µg/ml PfRh5 or BSG was prepared in PBS pH 7.5 and added to 96-well flat-bottomed plates (Maxisorp; Nunc) for overnight incubation at 4 °C. Plates were washed with PBS-T and then 200 µl of 5% w/v skim milk powder blocking solution made up in PBS-T was added, followed by incubation for 2.5-hours at room temperature, with mixing. Plates were then washed with 100 µl of each antibody hybridoma supernatants (supplied by Jenny Thompson (Walter and Eliza Hall Institute) for the anti-PfRh5 supernatants and from Dr. Anthony Hodder (Walter and Eliza Hall Institution) for the anti-BSG supernatants) were applied and incubated for one-hour at room temperature (the positive control polyclonal antibody was prepared in a 1: 10 000

blocking solution). Plates were washed, then incubated for two-hours at room temperature with 50 µl of a 1:2000 dilution of anti-mouse HRP. Development of the signals and analysis were carried out according to section 3.2.2, except 100 µl of acid was used for quenching. The BSG used was donated by Dr. Lin Chen. The PfRh5 used was from the previously described baculovirus batch production or an *E. coli*-produced version (Gennova) which was originally used to vaccinate the mice.

Protein G purification of antibodies was carried out in sterile conditions. Aliquots of 1.0 ml from each antibody hybridoma supernatant was mixed with 60 µl of a 50% protein G slurry (Qiagen) overnight at 4 °C. The sample was centrifuged at 1200 × g for one-minute, supernatant decanted, then two washes with cold PBS were carried out. After decanting the PBS, 45 µl of elution buffer was added (0.5 M NaCl, 0.1 M glycine, pH 2.6), mixed and centrifuged to recover the supernatant, which was added to tubes containing 5 µl of 1 M Tris pH 9.5 for neutralization. Samples were analysed via NanoDrop using a solution of 10% neutralization buffer and 90% glycine elution buffer.

For the G protein purified antibody screen, the same protocol as above was carried out but using the antibodies in a 1:100, 1:1000 and 1:10 000 dilution prepared in blocking solution.

The BSG inhibitory antibody ELISA screen was set up by coating 96-well flat-bottomed plates with 100 µl of 1 µg/ml PfRh5 made up in 1 × PBS, with an overnight incubation at 4 °C. A 1:2 dilution series of BSG was prepared across the lanes of a regular 96-well plate with a starting concentration of 10 µg/ml, whilst down each row various antibody dilutions were screened including 1:50, 1:100, 1:500, 1:1000, 1:2000, and 1:4000, in a final volume of 100 µl. The antibody + BSG mix was incubated at room temperature for two-hours. The rest of the

ELISA method was carried out as described above. The PfRh5 inhibitory mAb screen was carried out as described above but using EGF3-10 to coat the plate.

4.6.2 Results for PfRh5 and BSG Antibody Screens

The development of mAbs against PfRh5 and BSG was of interest for the identification of inhibitory mAbs acting against the receptor ligand interaction, in addition to the development of tools to be potentially used for crystal trials and various assays. After the generation of the hybridoma supernatants from PfRh5-vaccinated mice, screening for reactivity to PfRh5 was carried out (figure 4.48). The screen was carried out to compare results from the initial ELISA performed by Jenny Thompson using an *E. coli*-produced PfRh5, and the PfRh5 produced from baculovirus batches described in section 4.2. Relatively (taking into account differential maximum absorbance intensity), the results matched overall for the supernatants that gave the highest signals with the exception of sample 7E1. All samples in series 2 that showed an absorbance of over 3.5 were selected to be developed into mAbs. Several ELISA assays were set up to screen the supernatants for their ability to inhibit the PfRh5 – BSG interaction (results not shown); however, a successful assay remained elusive due to a perpetual problem of the control media giving high background noise and the positive control of BSG-coated plates with PfRh5 titrated on not giving a consistent and convincing enough signal.

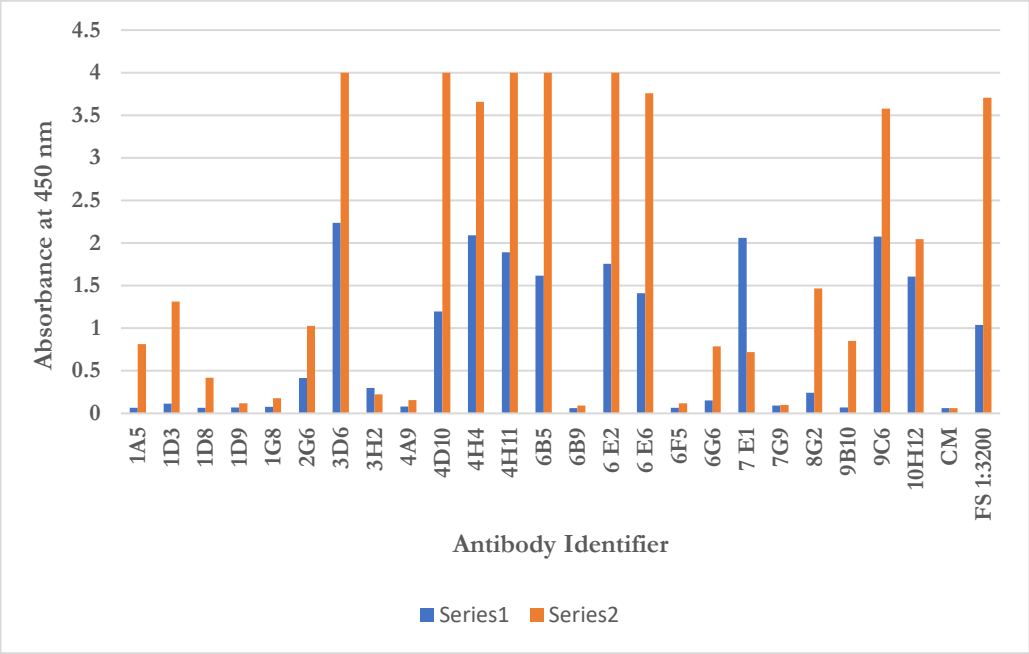


Figure 4.48. Initial Anti-PfRh5 Antibody Screen ELISA: Series 1- results of the antibody screen against *E. coli*-produced PfRh5 tested at the WEHI Antibody screening facility at Bundoora. Series 2- results from the antibody screen against the baculovirus-produced PfRh5. Abbreviations used: CM- control media (a negative control), FS 1:3200- fusion supernatant in a dilution of 1:3200 (a positive control).

After confirming the reactivity of hybridoma supernatants from BSG-vaccinated mice and concentrating the antibodies via protein G to around 0.06 mg/ml (results not shown), various dilutions of each antibody was tested for their reactivity to BSG (figure 4.49). Most mAbs were highly reactive, except for 4B5, 4C7, 4D3, 6E4, 5D11, 5E3, 6A5, 6B4, 4E4, 6F6, 7A3, and 8D12. Given that mAbs 8H3 and 5F7 showed the strongest reactivity against BSG, these were selected for follow-up studies to investigate their potential as inhibitory against the BSG – PfRh5 interaction.

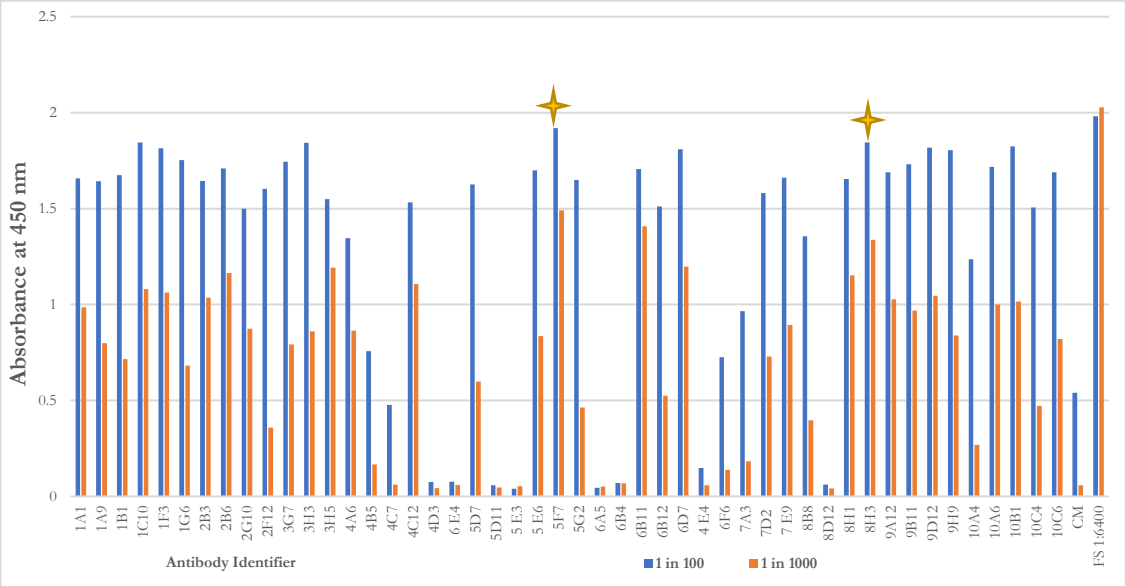


Figure 4.49. Anti-Basigin Antibody Reactivity ELISA: Results from the 1:10 000 dilution are not shown for both space consideration and the fact that this dilution reduced signal to negative control background level. Negative controls using 50 µl of 0.5 µg/ml BSG (detected with anti-mouse HRP) and PBS-only produced signals at 0.049 and 0.058, respectively. Yellow stars highlight the most effective antibodies.

A titration of BSG mixed with different dilutions of the F57 and 8H3 mAbs was applied to a PfRh5-coated ELISA plate to assess the effect of both BSG and mAb concentration (figure 4.50). Interestingly, as the concentration of BSG diminished, an increased signal was to be detected for both mAbs regardless of their dilution. This initially was suggestive that the mAbs were binding to the pFRh5 indirectly, however the controls in lanes 6 and 7 of figure 4.50 C do not reflect this. Clearly the experiment required repeating, however time and prioritization of other phases of the overall project did not allow for this.

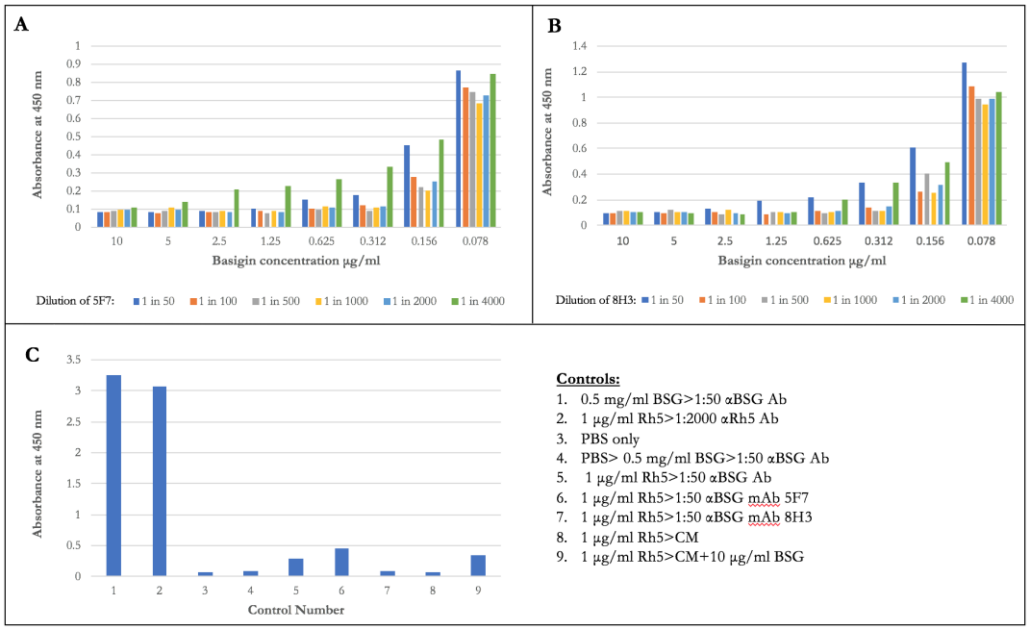


Figure 4.50. ELISA Screen of Anti-Basigin mAbs 5F7 and 8H3: A. Various concentrations of BSG that had been pre-incubated with various dilutions of the 5F7 anti-BSG mAb applied to 1 µg/ml of PfRh5 (then detected with anti-mouse HRP). B. Various concentrations of BSG that had been pre-incubated with various dilutions of the 8H3 anti- BSG mAb applied to 1 µg/ml of PfRh5 (then detected with anti-mouse HRP). C. Controls used. Each were detected with HRP-conjugated antibodies. Development time of lanes 1 and 2 took around 90-seconds before quenching (the rest of the controls were quenched after 30-minutes). Each absorbance reading given is the average of a triplicate. Abbreviations used: BSG- basigin, Ab- antibody, CM-control media.

4.7 MMV Malaria Box Compound Screen

4.7.2 MMV Malaria Box Compound Screen Methods

Approximately 1 mg of the PfRh5 produced in section 3.2 was used for screening against plates of the Medicines for Medicines for Malaria Venture (MMV) Box Compound Library (see https://www.ebi.ac.uk/chemblntd and Spangenberg, *et al.*, 2013) with assistance from A/Prof. Naomi Morrissette and Dr. Isabelle Lucet. A 1:100 stock dilution of each drug compound was used along with SYPRO Orange

dye (ThermoFisher) according to the general methods outlined in Lucet, *et al.*, 2014.

4.7.3 Results for the MMV Malaria Box Compound Screen

PfRh5 was screened against the MMV box in a TSA to look for any compounds that appeared to bind to the protein. Whilst many the compounds showed no evidence of this, or appeared to destabilise the protein, four potential hits were discovered (figure 4.51) after repeating their screening after an initial TSA was carried out. Originally the plan was to repeat these experiments with other dyes in addition to screening compounds produced by Dr. Brad Sleebs (of the Walter and Eliza Hall Institute); however, time and prioritization of other phases of the overall project did not allow for this.

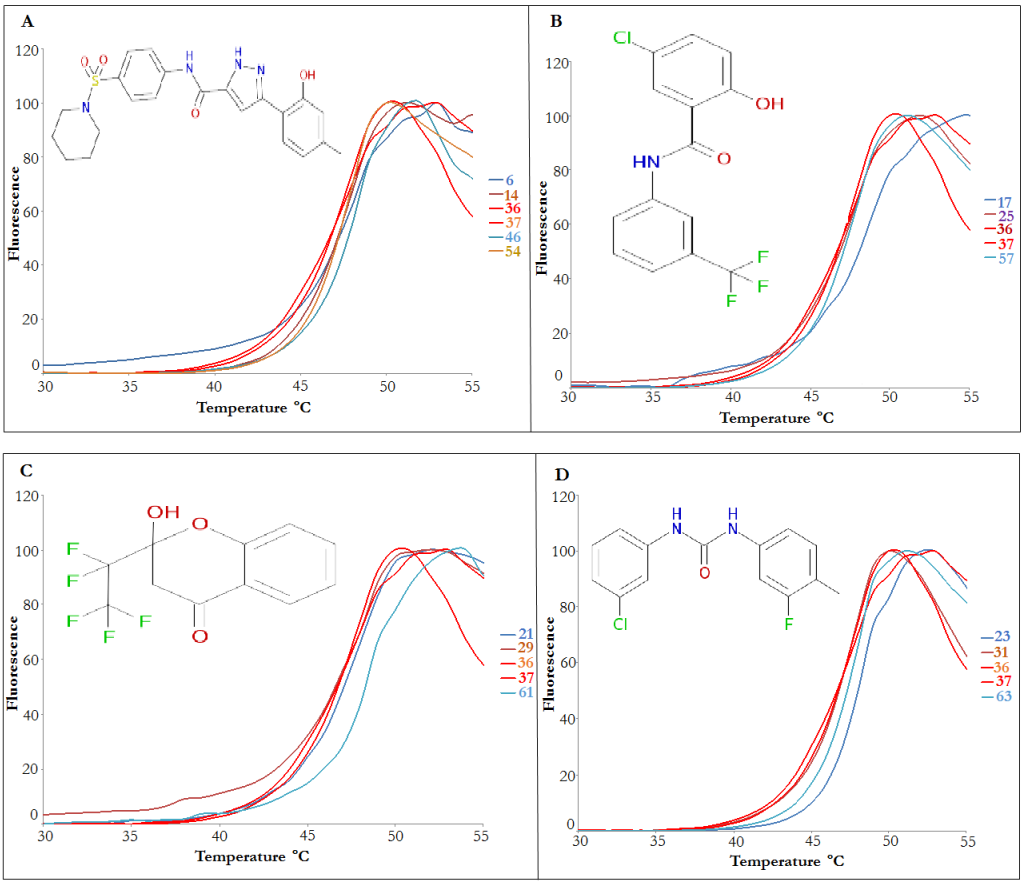


Figure 4.51. TSA Screen of PfRh5 and MMV Box Compounds: All screens show the chemical structure of each compound tested. A. Screen against compound MMV665827 (from MMV box B, well G10). When this compound was screened against *P. falciparum* parasites, the EC_{50} was 119 nM. 36 and 37 (red/orange) are the DMSO controls, 6 and 46 (blue) are repeats, 14 and 54 (brown) are 1:10 dilution repeats. B. Screen against MMV665807 (from MMV box B, well C11). 36 and 37 (red) are the DMSO controls, 17 and 57 (blue) are repeats, 25 (purple) is the 1:10 dilution. C. Screen against compound MMV000972 (from MMV box C, well E6). 36 and 37 (orange/red) are the DMSO controls, 21 and 61 (blue) are the repeats, 29 (brown) is the 1:10 dilution. D. Screen against compound MMV000911 (from MMV box C, well D7). 36 and 37 (orange/red) are the DMSO controls, 23 and 63 (blue) are repeats, 31 (brown) is the 1:10 dilution.

4.8 Discussion and Conclusions

Due to the necessity of PfRipr in the PAIN-complex and the dearth of information about this protein in contrast to the binding partners, it was important to develop tools for exploring this molecule. These tools first and foremost took the form of recombinantly-produced truncations of the full-length protein, with a focus predominantly on the EGF-like domains, given the role these domains commonly play in protein – protein interactions (Lin, *et al.*, 2001). The decision to concentrate on producing truncations of PfRipr was twofold; first, it might allow identification of a binding region for PfRh5. Second, their recombinant production would likely be far easier than trying to produce the large ~123 kDa full-length PfRipr protein, complete with its 87 cysteines that mostly form disulfide bonds. To this end, a set of seven EGF-rich proteins were produced in a soluble form in *E. coli*. The author's initial suggestion of utilizing an N-terminal polyhistidine tag, a maltose transporter subunit (MalE), maltose-binding protein (MBP) then a tobacco etch virus (TEV) enzymatic site before the N-terminal of each protein (MalE for transport to the periplasm in the *E. coli*, MBP for enhanced solubility and TEV for easy tag cleavage (Rosano & Ceccarelli, 2014)) was rebuffed in favor of a simpler design of each protein, whereby only the polyhistidine tag was kept. Fortunately, the simpler strategy worked for production of usable levels of recombinant protein. Initial purification was straightforward due to the presence of the N-terminal polyhistidine tag that allowed Ni-NTA affinity chromatography.

Four non-EGF truncated regions of PfRipr were also developed, including the middle section of the protein between the second and third EGF-like domains, of interest due to the yet unknown enzymatic cleavage point. Unfortunately, each of the non-EGF constructs demonstrated the need to be refolded, which at the time became a lower priority due to the decision to focus more exclusively on the EGF-rich regions. Returning to the EGF-rich regions, a standard purification protocol was developed to separate the multimers that were commonly seen after post Ni-affinity purification. This involved SEC, which typically resulted in a batch of pure,

monomeric protein, apart from EGF3-4, which over a short space of time, aggregated to make it unusable. As expected, each EGF-rich protein migrated faster on SDS-PAGE gels under NR conditions than R conditions, indicative of disulfide bonds being present which are one of the defining features of EGF domains.

The decision to move to an exploration of refolding some of the EGF-rich proteins was made due to the presence of highly concentrated and relatively pure protein in the pellet fraction (where insoluble inclusion bodies are found) when IPTG-induced at a high temperature, the protective effect against proteases when producing proteins as inclusion bodies (a significant amount of proteolytic degradation was seen in some samples previously), and due to some constructs having mixed batch to batch variation. Extensive refolding trials were performed to isolate the key conditions required to produce quality, monomeric and folded protein. Figures 4.11 and 4.12 show that over the three days of refolding, multimers began collapsing into mostly monomeric protein. To exploit differences in surface charge of monomer and resulting multimers, ion exchange chromatography was utilized (with SEC used to remove any contaminating dimer).

Once the proteins were refolded and purified, it was imperative to assess that they were the correct MW, had evidence of folding, and evidence of their predicted secondary structure (to assist in garnering confidence that misfolded protein had not been produced). Ideally, an activity assay would be carried out; however, since the PfRipr EGF-like regions had no evidence for enzymatic activity, nor a known binding partner, this was unable to be done. Therefore, following mass spectrometry, CD was utilized to reveal an elliptical signature of EGFs 5-6 and 3-6 highly similar to another β-sheet containing *P. falciparum* protein comprised also of tandem EGF domains. With respect to EGF5-6, a more quantitative analysis of the CD spectrum revealed that, as expected, the majority of the protein was likely disordered with ~22% being β-strand. This was reasonable evidence that the protein was not only folded, but contained the amount and type of secondary

structure typical of an EGF-like domain. Further evidence for folding of EGFs 5-6 and 3-6 came from the melting curve generated by the TSA (figure 4.23) since if the proteins were unfolded, they would generally bind dye molecules at low temperatures, leading to an increased fluorescence signal that would not change significantly as the temperature was raised.

To assess the presence of any free cysteines in EGFs 5-6, 7-10 and 3-10, a PEGylation experiment was carried out whereby the maleimide would bind any reduced cysteines, and the attached PEG group would slow the electrophoretic mobility of the protein during SDS-PAGE, presenting as a defined, higher MW band (Lu & Deutsch, 2001). As shown in figure 4.24, around 95% of the protein remains in the original position on the gel, suggesting that only a small proportion of the sample has free cysteines (indicative of the protein being misfolded or partially unfolded), which is not uncommon in a sample of refolded recombinant protein (as mentioned and reviewed in Singh & Panda, 2005).

Finally, the use of NMR was important for assessing the refolding of proteins since one can make use of 1D and 2D NMR to see strong evidence of protein folding by looking for good peak dispersion (1D and HSQC), and evidence of β-sheet. Figures 4.26, 4.28 and 4.29 demonstrate both good peak dispersal in the carbon and amide regions, in addition to presenting evidence for secondary structure in the form of alpha protons in β-sheet at a chemical shift of around 5 ppm. Overall, and as is common for proteins undergoing NMR analysis, a higher temperature and lower pH yielded better quality spectra with respect to sharper peaks and better resolution. As the initial plan was to pursue NMR-based structural studies on the EGFs that gave rise to the best spectra, a natural abundance HSQC and NOESY were carried out on EGF5-6 and 7-10, with the spectra from EGF5-6 yielding the clearest spectra for each experiment. The ^{1}H-^{1}H NOESY of EGF5-6 additionally displayed evidence of folding due to the presence of many cross peaks between key regions such as long-range methyl protons coupling with aromatic protons, amide – amide protons, and αH to amide protons.

The excellent-looking 2D NMR experiments for EGF5-6 thus indicated it may be an ideal protein on which to pursue for NMR-based structural studies. This would have been pursued further if the decision to move into a *Pichia*-based recombinant protein production system had not been made.

It should be noted that numerous attempts at crystallization were attempted but ultimately failed to yield protein crystals. The failure of these crystallization attempts was potentially due to the proteins containing highly flexible regions. This can lead to thermodynamically unfavourable conditions for reaching the energy minima needed to facilitate a protein crystalizing. Further, at the time of experiments, no mAbs against PfRipr had been developed which could have facilitated successful crystallography. Protein quality is also another consideration that may have impeded successful crystallization attempts.

The significance of refolding (and subsequent biophysical characterization of) these recombinantly-produced EGF-rich proteins that has been demonstrated must not be underestimated; proteins are both complex and dynamic entities. Bacterial expression systems can be effective at rapid production of proteins in large quantities, but can be poor at forming native disulfide bonds, which are inconsistent with producing a native protein conformation. Furthermore, the paucity of PfRipr biochemical publications making use of recombinant protein production (at the time of the original research outlined in this thesis) are highly suggestive that other scientists have had difficulty in both producing and convincingly validating their PfRipr proteins. The work in this chapter therefore represents the first time recombinant PfRipr proteins have been produced with biophysical characterization. Hence, the methods outlined in this thesis should be of interest and use to other malariologists and protein biochemists in particular.

With respect to the PfRh5 – PfRipr interaction studies, these become problematic due to a variety of factors, including the "stickiness" of the recombinant PfRh5 with the bead matrix used in coIP assays and in ELISAs. The initial ELISA in

figure 4.35 appeared to show convincing results of an interaction between PfRh5 and either EGF5-6 or EGF3-6, due to a dose-response curve that was being searched for. However, the lack of ability to repeat the results convincingly was problematic and led to two possibilities; either the initial result was correct and perpetual human error occurred in the rest of the repeats, or, applying Occam's razor and the knowledge of the focusing and anchoring effect in bias, all the negative results were in fact real. As of February 2017, the scientific community now knows that PfRh5 and PfRipr do not in fact form a tight complex (Galaway, *et al.*, 2017). However, this work was carried out between 2013 and 2016, and due to the possibility that either the fifth and/or sixth EGF-like domain were indeed responsible for binding to PfRh5 (or at least, elsewhere on PfRipr), a variety of other assays were pursued.

After moving away from coIP assays, FRET analysis was attempted but failed likely due to one or both of the fluorophores binding at or near the binding region between BSG and PfRh5, thus preventing a positive control being established. The BN-PAGE results (data not shown) were suggestive that no binding event was taking place between the PfRipr EGF-like domains and PfRh5. However, since the positive control was not shown to conclusively work, these results remained inconclusive. It was therefore decided to see if a tight complex was formed between either PfRh5 and EGF3-6 or EGF3-10 which could be detected via SEC/gel filtration. Due to the potential of weaker protein complexes destabilizing during their permeation and movement through the resin in a SEC column, the method is only able to inform one on if a tight complex is formed between two biomacromolecules. In the case of PfRh5 and the PfRipr proteins, this was conclusively ruled out, thus representing the first evidence that PfRipr does not in fact associate with PfRh5, in agreeance with the published work that came out later by Galaway, *et al.*, 2017. Unfortunately for the author and the scientific community at large, publishing negative data on its own is still discouraged in academia (and is spoken about more than it is actually done (Young, *et al.,* 2008) (Collins & Tabak, 2014)).

In the future, it would be of interest to produce a batch of the PfRipr EGF-rich constructs, divide the batch in two; one of which would have the polyhistidine tag removed (with TEV, given that EK cleavage was unsuccessful with pET45b+ plasmid produced samples). Then one would be able to conduct an EGF – EGF interaction ELISA to screen using anti-pentahistidine antibodies. Because it may in fact be the case that the EGF domains of PfRipr are binding to one another to hold the two large post-cleaved polypeptides of the protein together in the PAIN-complex.

The development of both inhibitory compounds against the formation of the PAIN-complex proteins and tools to further probe said complex was of interest. If mAbs could be generated they could be utilized in crystallography studies (for example, by helping stabilize a protein or proteins), or epitope mapping studies similar to those in Lim, *et al.*, 2015 and as described in section 3.2. An initial selection of eight hybridoma supernatants containing anti-PfRh5 antibodies was made following an ELISA to assess their reactivity against PfRh5. It was however a shame that both the anti-PfRh5 antibody supernatants and the anti-BSG mAbs gave inconclusive results in the ELISA assays to assess their potential of inhibiting the interaction between BSG and PfRh5. This was the result of a high background signal to noise ratio produced by indirect interaction of the PfRh5, and the antibodies with the ELISA assay. It was decided to provide aliquots of these antibodies to Drs. Wai-Hong Tham and Anthony Hodder to conduct alternative assays, such as the FACS-based assay described in Lim, *et al.*, 2015.

With respect to the screening of the MMV box compounds against PfRh5 via the TSA, what one is looking for is initial evidence of binding between the small molecule and the protein. This is demonstrated when (compared to the control with DMSO instead of the compound of interest) there is a protein melting curve shift to the right, indicating that more energy is required to denature the protein, which has become somewhat stabilized due to the compound binding. The four

compounds identified in figure 4.51 display this to varying degrees. Of course, the experiment would need to be repeated with the use of an alternative dye to make sure that was not a contributing factor, then the compounds would need to be tested in a parasite invasion assay. Further, some of the compounds may require some medicinal chemistry to make them more "druggable". However, in favour of these molecules is that none of them violate any of Lipinski's rule of 5 (Lipinski, 1997). Though, MMV665827 and MMV665807 violate one of the rules of 3, namely that they are over 300 Da and the former compound has over three hydrogen bond donors/acceptors (although, the rule of 3 is more important for lead compounds) (Congreve, *et al.*, 2003).

Due to significant time constraints, it was decided to focus exclusively on the production of PfRipr proteins in a *P. pastoris* system, using a proprietary strain developed by Dr. Andrew Herbert in the Barlow lab at the University of Edinburgh to avoid the need to work with refolding, and to ensure disulfide bond formation. The direction of the project was therefore altered to be focused on new protein production/purification, crystal trial screens of larger PfRipr constructs, additional biophysical characterization, binding studies between PfRipr proteins and PfRh5 or CyRPA, assessment of calcium binding in PfRipr, and extensive NMR studies.

Chapter 5

Investigations into *Pichia*-produced PfRipr

"*The struggle itself towards the heights is enough to fill a man's heart. One must imagine Sisyphus happy*".

Albert Camus, *The Myth of Sisyphus and Other Essays* 1942.

5.1 Overview

The choice to move into a *P. pastoris*-based protein production system was made primarily due to the ability to use a proprietary strain (dubbed the "superstrain" (SS) *Pichia*) that had been genetically-engineered by members of the Barlow group at the University of Edinburgh. This had been designed specifically for the production of disulfide-rich proteins that are exported into the culture media at potentially high levels. This would avoid costly protein refolding protocols as required in *E. coli* and provide the possibility of producing larger PfRipr constructs, including the full-length protein.

Compared to *E. coli*, *P. pastoris* has other benefits as a eukaryotic expression system, including the ability to glycosylate proteins (however, this can be disadvantageous as glycans added are dissimilar to mammalian glycans), other post-translational modifications, and correct proteolytic processing (Cregg, *et al.* 2000) (Cereghino & Cregg, 2000). Furthermore, *P. pastoris* is highly amenable to large-scale fermentation, allowing for economically viable production of recombinant proteins of high quality and in large quantities (Cereghino, *et al.* 2002). This is particularly interesting to those wanting to develop industrially scalable methods of producing protein-based vaccine components. *P. pastoris* belongs to a group of methylotrophic yeasts, i.e. they use methanol as their sole carbon source. The metabolism of methanol involves several steps, starting with alcohol oxidase (AOX) which oxidises the methanol into formaldehyde and hydrogen peroxide. The latter product is a harmful by-product that is isolated in the peroxisome and processed into gaseous oxygen and H_2O via the activity of catalase. Given that AOX has a low affinity for O_2 and is produced in high quantities by the cell, the promoters concerned are particularly strong, and hence effective for use in producing high levels of recombinant protein.

P. pastoris utilises two AOX genes; *aox1* and *aox2*. The former is responsible for more effective metabolism of methanol and therefore overall cell growth, whilst

being regulated at the transcriptional level. This regulation involves repression, de-repression or induction. With respect to repression, when glucose is present, the activity of *aox1* is significantly reduced even when methanol is also present. Therefore, use of a glucose-rich medium provides a convenient way of growing cells to a high density initially. When induction of AOX1 is required for recombinant gene translation, one simply swaps out the glucose-rich media for one containing methanol. There are two phenotypes of *P. pastoris* that utilise methanol; MUT^S and MUT^+. The former lacks *aox1* (and therefore lacks most AOX activity) and relies on AOX2 for slow growth with methanol. The latter, MUT^+, grows on methanol-only sources like their wild-type counterparts.

The base-strain type of *Pichia* used in this study was SuperMan5, belonging to the MUT^S phenotype. However, for some initial trial experiments the KM71 (of the MUT^+ phenotype) base-strain was used. To exploit the AOX1 promoter, the plasmid pPICZαβ was used as a vector. This vector possesses DNA coding for the α-factor secretion-signal peptide which is fused to the inserted gene of interest to ensure the recombinant protein product is secreted. For removal of the α-factor signal sequence, the *kex2* gene product cleaves between the Arg and Gln in the sequence- Glu-Lys-Arg|Glu-Ala-Glu-Ala (Brake, *et al.* 1984). This is followed by STE13-mediated cleavage of the Glu-Ala repeats; though this can be inefficient, leaving N-terminal non-native repeats on the protein of interest (Julius, *et al.* 1983).

This chapter overall aims to show the production of EGF-rich domains of PfRipr in the aforementioned *P. pastoris* system, with varying degrees of success. Only the proteins that showed definitive signs of stable expression were pursued for further development, including biophysical characterization and structural studies. Of those that were successful, one (EGF5-7) could be taken to the stage of double and triple resonance NMR experiments using a $^{15}N/^{13}C$-labeled protein that proved to be stable over time and a range of temperatures. This yielded data that should lead to a 3D molecular structure (personal communication with Dr. Brian Smith, June 2017).

Small angle X-ray scattering (SAXS) coupled to homology replacement modelling was used to generate a 3D structural model of EGF5-7, that will also be of use in the future for validation of NMR-based structural modelling. SAXS is a structural technique for monochromatix X-ray beam analysis of a sample in solution. The scattered intensity [I] is recorded as a function of the scattering vector 223 $s=4\pi\sin\theta/\lambda$ where 2θ is the angle between the incident and the scattered beam. Scattering generated from the solvent is measured separately and then subtracted to give the final data set. This background-corrected intensity [I](s) is proportional to the scattering of a particle averaged over each molecular orientation. Various 1-D curves can be extracted from the [I](s) information, providing important parameters directly related to the molecule's size, oligomeric state and overall shape. Analysis of the scattering patterns allows for *ab initio* molecular shape reconstruction and rigid body modelling in addition to analysis of potential molecular flexibility. In lieu of a crystal or NMR-based structure, SAXS can be utilized for a level of validation of homology-based molecular modelling and the confirmation of molecular rigidity that one gains from relaxation data generated from NMR experimentation. Further, SAXS provides useful information that can be used to refine NMR-based structural data.

In addition to garnering structural information, it was imperative to screen proteins for binding to CyRPA and to assess the potential of calcium cation binding due to the importance of Ca^{2+} activity demonstrated by the PAIN-complex (Volz, *et al.* 2016). Finally, the development and screening of inhibitory mAbs against PfRipr was of importance for validation of recombinant PfRipr as a vaccine candidate, and for the identification of where in PfRipr inhibitory mAbs bind. This would form the first steps towards epitope mapping. The list of where each protein got to in the examination phase is shown in table 2.

Protein	Purified	MS	CD	TSA	Free Cys Check	1D and 2D NMR	3D NMR	Crystal Trial	SAXS	Rh5 Binding Assay	CyRPA Binding Assay	WB
EGF1-2	N	N	N	N	N	N	N	N	N	N	N	N
EGF3-5	N	N	N	N	N	N	N	N	N	N	N	N
EGF5-7	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
EGF7-10	Y	Y	N	N	N	Y	N	Y	N	N	Y	Y
EGF3-10	N	N	N	N	N	N	N	N	N	N	N	N
Ripr21-1086	N	N	N	N	N	N	N	N	N	N	N	Y
Ripr21-604	N	N	N	N	N	N	N	N	N	N	N	N
Ripr604-1086	N	N	N	N	N	N	N	N	N	N	N	N

Table 2. PfRipr Protein Progress Log (SS Pichia): The yellow highlighted regions indicated where a particular examination type was carried out.

5.2 Cloning, Test Expression and Purification of PfRipr Constructs in *Pichia pastoris*

5.2.1 Methods for the Cloning, Test Expression and Purification of PfRipr Constructs in *Pichia pastoris*

LB agar plates containing 500 μg/ml zeocin (500Z) of freshly re-streaked KM71 and SuperMan5 (engineered to produce proteins with $Man_5[GlcNAc]_2$ Asn-linked glycans) SS Pichia were available in the Barlow group. Synthetic genes subcloned into pPICZαβ plasmids were expression-optimised and provided by Geneart (ThermoFisher). They included C-terminal hexahistidine tagged PfRipr from amino acids 21 – 1086 (Ripr21-1086), 21 – 604 (Ripr21-604) and 604 – 1086 (Ripr604-1086), in addition to non-tagged (due to an engineered-in stop codon at the C-terminus) EGFs 1-2, 3-5, 5-7, 7-10 and 3-10.

All plasmids from Geneart were resuspended in 50 µl of ultrapure H_2O (generating a stock of 0.1 µg/µl plasmids), with 1 µl pipetted into 49 µl of ultrapure H_2O (generating the usable stock at a concentration of 2 ng/µl). 1 µl of the stable stock was added to a 50 µl tube of competent Top10 *E. coli* cells (ThermoFisher), incubated on ice for 30-minutes then heat-shocked at 42 °C for 30-seconds. Tubes were placed back on ice for one-minute, 0.25 ml SOC medium added then incubated at 37 °C for one-hour. Aliquots of 20 and 100 µl of the culture was plated onto 25Z low-salt LB agar plates and incubated at 37 °C overnight. Maxipreps were carried out on picked colonies as per manufacturer instructions (Qiagen CompactPrep Plasmid Maxi Kit).

Restriction enzyme digests were carried out using 0.5 ml of plasmid DNA from maxipreps, 5 µl *Sac*I-Hf and 45 µl of CutSmart® buffer (New England Biolabs). The mix was incubated at 37 °C for one-hour, then analysed on a 1% or 1.2% agar TAE gel with 0.01% Sybr Safe added (ThermoFisher), with 15 µl of a 1 kb DNA ladder (New England Biolabs) added to lane 1, and 1.5 µl of digested plasmid in 10 µl of 1 x loading dye distributed in the other lanes. The gel was run at 100 mA for 45-minutes. Phenol-chloroform extraction and ethanol precipitation was carried out as per section 2.1.

Pichia transformations were carried out with Roma Galloway and Ella Svahn according to the protocol outlined in section 2.1.15.3 (Makou, 2013) using a Bio-Rad Gene Pulser for the elecroporations. Once visible colonies were seen they were restreaked onto 300 then 500Z yeast extract peptone dextrose (YPD) media.

Test expression was carried out with Ella Svahn, according to the general protocol outlined in section 2.1.15.5 of (Makou, 2013), with the following changes- 100 ml of buffered complex medium containing glycerol (BMGY) used for starter cultures, and 25 ml of buffered complex medium containing methanol used for culture resuspension. Recipes for the standard media used were found in the

EasySelect™ Pichia Expression Kit 2010 (Invitrogen). Glycerol stocks were prepared according to the protocol in section 2.3.

For the prediction of potential glycosylation sites, the NetNGlyc 1.0 Server was used (http://www.cbs.dtu.dk/services/NetNGlyc/). For deglycosylation tests, 1 - 5 µl of *Endo*-Hf or *Endo*-Hr (New England Biolabs) was added to each post-harvest 20 µl supernatant sample for up to one-hour at 37 °C. This enzyme was used for removing mannose-rich N-linked carbohydrate moieties from the recombinant proteins by cutting between the two first GlcNAc molecules thus leaving the protein with one GlcNAc moiety per glycosylation site.

SDS-PAGE and Western blot analysis was carried out with occasional assistance from Ella Svahn, as described in section 2.8 using 12.5% acrylamide NuPAGE Novex gels or Bolt Bis-Tris Plus gels (Invitrogen), in 10% NuPAGE 20x MES or MOPS buffer and electrophoresed at 185 V for 45 – 55-minutes using an X-cell Sure Lock apparatus (Invitrogen). Western blots made use of rabbit polyclonal anti-PfRipr antibodies (1:2000 dilution) and goat anti-rabbit HRP (1:1000 dilution), with development either via nitrocellulose and X-ray film or 3,3`-diaminobenzidine tablets (DAB; Sigma-Aldrich) on a methanol-activated PVDF membrane (GE Life Sciences).

Ni-affinity purification was carried out using 50 µl of Ni-NTA beads (Roche) per 1 ml of *Pichia* supernatant, with a one-hour incubation on a rotor at room temperature. Samples were centrifuged at 1200 × g for one-minute, supernatant carefully removed via pipette and then three 0.2 ml washes with wash buffer were carried out, followed by elution with 100 µl of elution buffer. A 35 µl sample was analysed via SDS-PAGE and Western blot using methods described in section 2.4 and 2.8 (respectively).

Cation exchange chromatography was carried out as per the general methods outlined in section 2.2.6 of (Makou, 2013). Buffer A contained 20 mM sodium acetate (at either pH 4 or 6) and buffer B containing the same as A but with the addition of 1 M NaCl. Aliquots of 20 ml of each supernatant were diluted with ddH_2O by a factor of 1:1, 1:2, 1:3, or 1:5. 5 ml elutions were collected, and of those 20 µl was analysed via SDS-PAGE under reducing conditions. Ella Svahn provided occasional assistance with the chromatography of EGF5-7.

5.2.2 Results of the Cloning, Test Expression and Purification of PfRipr Constructs in *Pichia pastoris*

After transformation, plating and incubation of Top10 *E. coli* with the various plasmids, it was found that Ripr21-604 gave rise to 250 colonies on the 100Z plate, Ripr604-1086 gave rise to ~600 colonies, Ripr21-1086 gave rise to 21 colonies, EGF3-10 gave rise to 100 colonies, EGFs 1-2, 3-5, 5-7 and 7-10 gave rise to thousands of colonies on each plate. Around 130 µg of DNA was obtained per maxiprep (in a volume of 0.55 ml), and ran according to their expected sizes once the purified plasmids were linearized with *Sac*I-Hf and analysed by agarose gel electrophoresis (data not shown). Following the *P. pastoris* transformation, clear individual colonies appeared on each plate (data not shown), with those on the 500Z plates used for test expression and resulting analysis via SDS-PAGE (figure 5.1). Both EGF5-7 and 7-10 yielded up to ~1 mg/ml for some colonies screened. As EGF7-10 was predicted to contain a single possible glycosylation site, a replica gel of each sample was run after the supernatant had been incubated with *Endo*-Hf; however, no difference was visualized upon examination indicating that the predicted site may have been in an inaccessible location within the molecule.

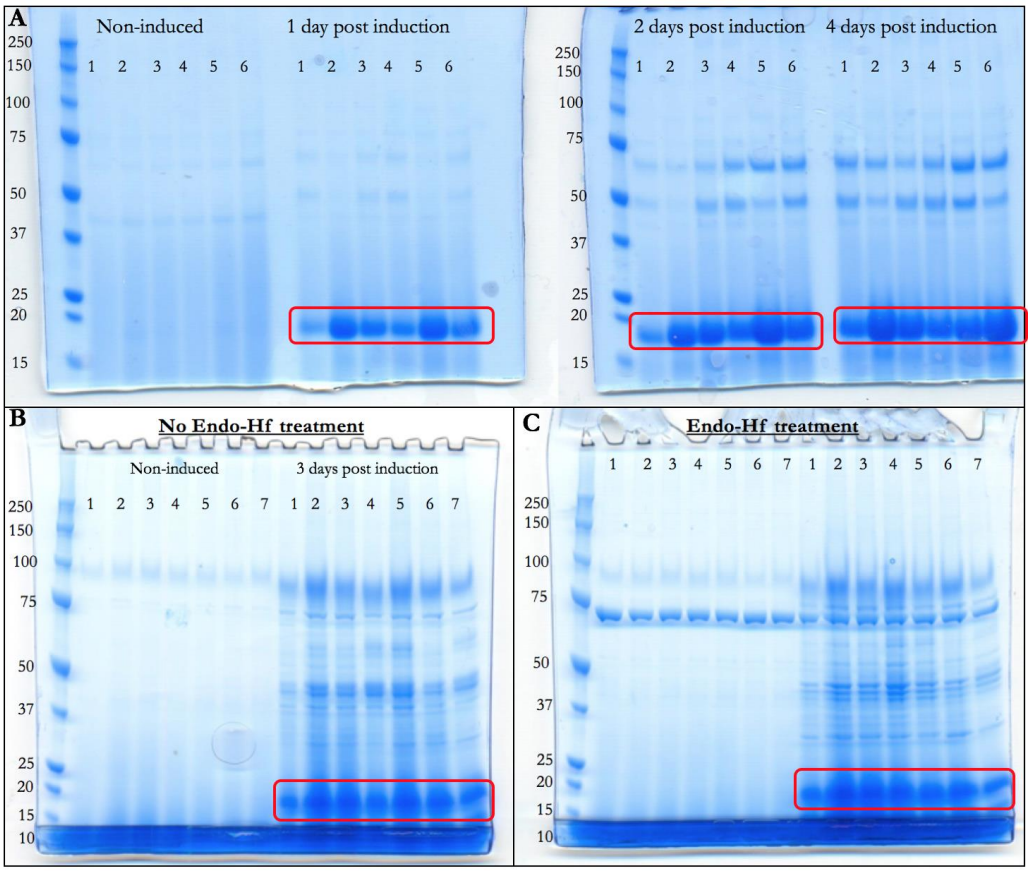


Figure 5.1. Test Expression of *Pichia*-Produced PfRipr Proteins: A. EGF5-7 cultures pre- and post-induction over four days with each culture from a separate colony from individual plates (numbers 1 – 6) screened via SDS-PAGE. B. EGF7-10 cultures pre- and post-four days of induction with each culture from a separate colony from individual plates (numbers 1 – 7) screened via SDS-PAGE. C. To assess if *Endo*-Hf made a difference to the EGF7-10 bands, another SDS-PAGE gel was prepared as per B, but with 1 µl of *Endo*-Hf enzyme added to each sample. 10 µl of supernatant added to each lane for gels A - C, expected sizes are highlighted in the red box.

With respect to EGFs 1-2, 3-5, 3-10 and Ripr21-604, Ripr604-1086 and Ripr21-1086, no bands were visualized on an SDS-PAGE gel, even following a ten-fold Ni-NTA-based concentration was attempted on each Ripr protein containing a C-terminal hexahistidine tag (data not shown). A Western blot was carried out on each of these samples in case the protein was being expressed in very low levels. Only Ripr21-1086 showed bands indicative of protein production (figure 5.2 A), with all induced samples giving a band migrating around 175 – 200 kDa (the

expected size was 130 kDa, however 11 glycosylation sites were predicted). Figure 5.2 B and C show the presence of EGFs 5-7 and 7-10. Unfortunately, the original plate with the Ripr21-1086 colonies had become contaminated and became unusable. After numerous re-transformations to generate more *Pichia* colonies to screen, none turned up any evidence of protein expression (Ripr21-604 initially appeared to show expression, but mass spectrometry analysis (data not shown) revealed the protein to be Factor H, therefore the KM71 *Pichia* I was given had been the wrong type), so after 18 months of various attempts, they were dropped from the project so as to focus on the *Pichia*-produced proteins that looked promising (especially given that there were still the *E. coli*-based PfRipr proteins to be used).

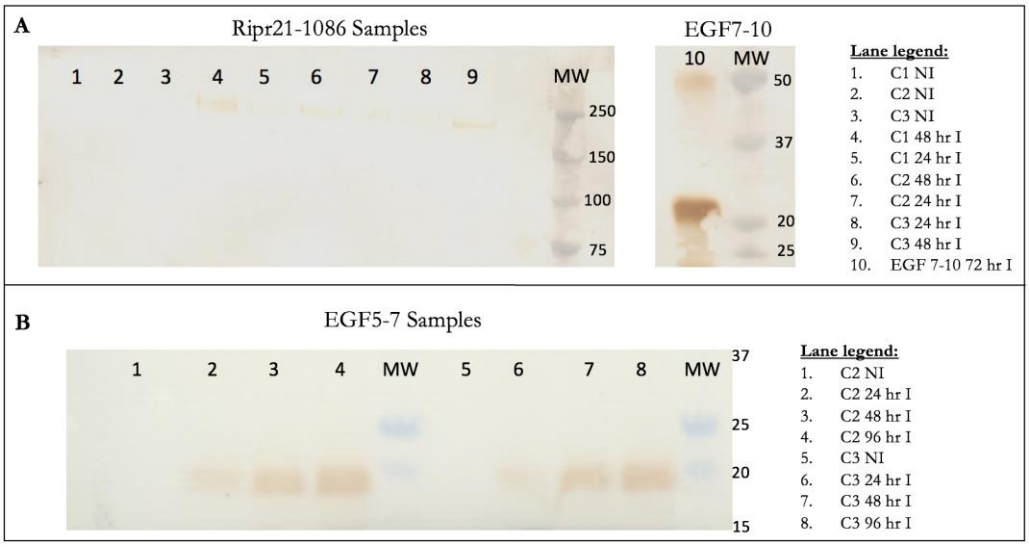


Figure 5.2. Western Blot Detection of Ripr21-1086: A. Detection of Ripr21-1086 from three colonies and EGF7-10 from a single colony. B. Detection of EGF5-7 from two colonies. All samples run under reducing conditions. Rabbit polyclonal anti-PfRipr antibodies used for detection, and DAB for development.

Colony 2 for SS *Pichia* strains expressing EGFs 5-7 and 7-10 was maintained via monthly re-streaking and glycerol stocks prepared. To optimize cation-exchange chromatography of these proteins, cell culture supernatant-dilution factor and pH tests were performed. An example of the results is given in figure 5.3 (EGF5-7

results looked the same, data not shown). There was a demonstrable increase in protein bound then eluted when the dilution was increased. Varying the pH had a negligible effect between 5 and 6, so going forwards it was decided to use the buffer at pH 4 and a dilution of 1:5. This lower pH would also be beneficial for NMR-based studies.

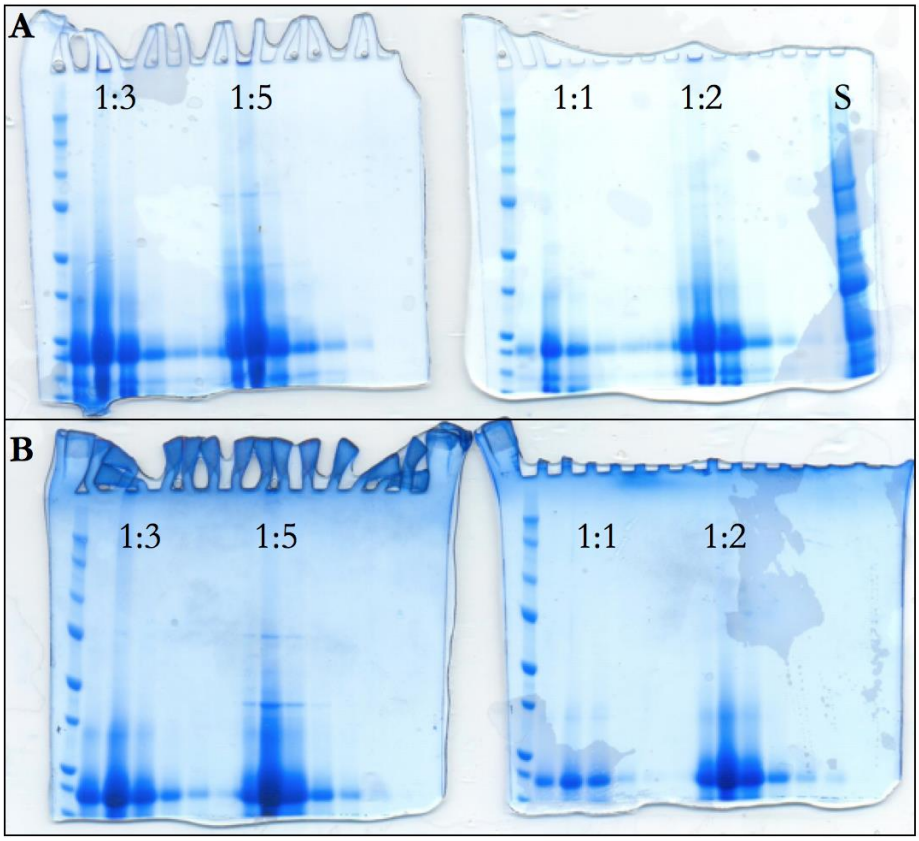


Figure 5.3. SDS-PAGE Analysis of Test Cation Exchange Conditions for EGF7-10: A. Samples from five 2 ml elutions at various dilution factors using a sodium acetate/NaCl buffer at pH 6. S- supernatant, loaded neat. B. Samples from five 2 ml elutions at various dilution factors using a sodium acetate/NaCl buffer at pH 4.

5.3 Scale-Up Production and Purification of Lead EGF-Rich Modules

5.3.1 Methods for the Scale-Up Production and Purification of Lead EGF-Rich Modules

Scale-up protein production was performed as per protocols listed in sections 2.1.15.6 and 2.1.15.7 (with the change of using 10 L culture media in a 12 L instead of a 3-L fermenter, carried out by Dr. John White) in Makou, 2013. SDS-PAGE and Western blot analysis was carried out as described in section 4.2.1 of Makou, 2013. Where protein concentration was required, pre-washed and equilibrated 5 kDa cut-off protein concentration tubes (Merck Amicon Ultra-15 Centrifugal Filter Units) were used in a chilled centrifuge, spinning for 20-minute periods at 3500 × g until the desired volume was reached. Protein concentration was determined via NanoDrop as described in section 2.7.

For producing ^{15}N-labeled proteins in SS *Pichia,* a two-day starter culture (10 ml) of BMG media (100 mM potassium phosphate pH 6, 1.34% YNB, 4 × 10^{-5}% biotin, 2 g/L ^{15}N-ammonium sulfate, and 1% glycerol) was grown in a 30 °C incubator at 220 rpm, then transferred to 1-L of BMG media until the cell growth was around an OD_{600} of 14 – 18. At this point the cells were centrifuged at 3500 × g for 30-minutes at 4 °C and resuspended in 250 ml BMM (with 0.5% methanol and 0.5 g ^{15}N ammonium sulfate), then incubated at 18 °C. Twice daily feeds were carried out with 1.25 ml methanol for three days, with 0.1 mM Pefabloc® (Sigma Aldrich) added on day 3. Final cell harvest (day 4) to obtain supernatant was via centrifugation at 10 000 × g for 30-minutes at 4 °C (then repeated).

For producing a double-labelled (^{15}N/^{13}C) protein sample of EGF5-7 in SS *Pichia,* a two-day starter culture (10 ml) of BM media (100 mM potassium phosphate pH 6, 1.34% YNB, 4 × 10^{-5}% biotin, 2 g/L ^{15}N-ammonium sulfate, and 5 g/L ^{13}C-U-D-glucose) was grown in a 30 °C incubator at 220 rpm, then transferred to 1-L of BM media containing 0.5 ml antifoam agent 204 (Sigma Aldrich) until the cell

growth is arrested. At this stage 1 g of ^{13}C-U-glycerol was added which allowed the cells to reach an OD_{600} of 14. At this point the cells were centrifuged at 3500 × g for 30-minutes at 4 °C and resuspended in 250 ml BMM (with 0.25 ml antifoam agent 204, 1.25 ml ^{13}C methanol and 0.5 g ^{15}N ammonium sulfate), then incubated at 18 °C. Twice daily feeds were carried out with 1.25 ml ^{13}C methanol for four days. Final cell harvest (day 4) to obtain supernatant was via centrifugation at 10 000 × g for 45-minutes at 4 °C (then repeated). The double-labelled-like (dll) sample was made as above but with non-isotopically labelled carbon sources.

For cation exchange chromatography of unlabelled samples, diluted supernatant had the pH checked and adjusted to pH 4 before being slowly pumped onto a 38.5 ml XK 26 column preparative column (GE Life Sciences) containing negatively charged SP-Sepharose (in the cold room), then washed (buffer A, 20 mM sodium acetate pH 4) and eluted using a gradient up to 100% of buffer B (20 mM sodium acetate pH 4 and 1 M NaCl) on an ÄKTA-Pure 25 FPLC system (pump P-920, UV detector unit UPC-900) using Unicorn 7.1 control software at the EPPF. Fractions of 15 ml were collected during a flow rate of 3.5 ml/minute. For isotopically labelled samples, a 5 ml FF SP cation exchange column (GE Healthcare) was used, collecting 0.5 ml fractions at a flow rate of 10 ml/minute.

Size-exclusion chromatography was performed as a subsequent purification step using a HiLoad 26/60 Superdex 75 preparative grade (GE Healthcare). Prior to loading the column was equilibrated with 1.1-1.6 column volumes of the appropriate buffer (typically 20 mM sodium acetate pH 4 and 100 mM NaCl). Elutions were in 1.5 ml volumes collected at a flow rate of 2.5 ml/minute. Where required, a HiLoad 16/60 Superdex 75 column was used at a flow rate of 1 ml/minute, collecting 1 ml fractions.

Where further purification was required, either MonoQ 5/50 GL cation chromatography (using buffer A and B) with a 1 ml column was used following

tenfold sample dilution and collecting 0.25 ml fractions, or reverse phase high performance liquid chromatography (RP-HPLC) was carried out collecting 0.5 ml fractions (as carried out by Dr. Martin Wear at the EPPF). With respect to the RP-HPLC, 100 µl of protein was loaded onto a Source15 RPC 4.6/100 ST column that was pre-equilibrated in 0.065% trifluoroacetic acid (TFA) and 8% acetonitrile (ACN). A 20 – column volume gradient (1.6 ml column volume) was used with the gradient from 0.065% TFA + 8% ACN to 0.05% TFA + 80% ACN.

Ella Svahn provided occasional assistance for the production of all forms of EGF5-7.

5.3.2 Results for the Scale-Up Production and Purification of Lead EGF-Rich Modules

Following on from the successful test expression and purification of EGF5-7 and 7-10, a 10-L fermentation was carried out to generate a total of ~10 g of each SS *Pichia*-produced protein. The resultant supernatants were purified via cation exchange chromatography with SDS-PAGE analysis of the key fractions (figure 5.4). EGF5-7 eluted at around 15 – 20% buffer B (or, 150 – 200 mM NaCl), with 5 ml each of fractions A9 – B2 pooled and concentrated to 6 ml. The results shown for EGF7-10 are from the flow-through of the initial batch binding to the column being put back over and eluted (hence the lower mAU values compared to EGF5-7). Here, EGF7-10 is shown to elute at around 20 – 25% buffer B, with 5 ml each of fractions A12 – B8 combined and concentrated to 9 ml.

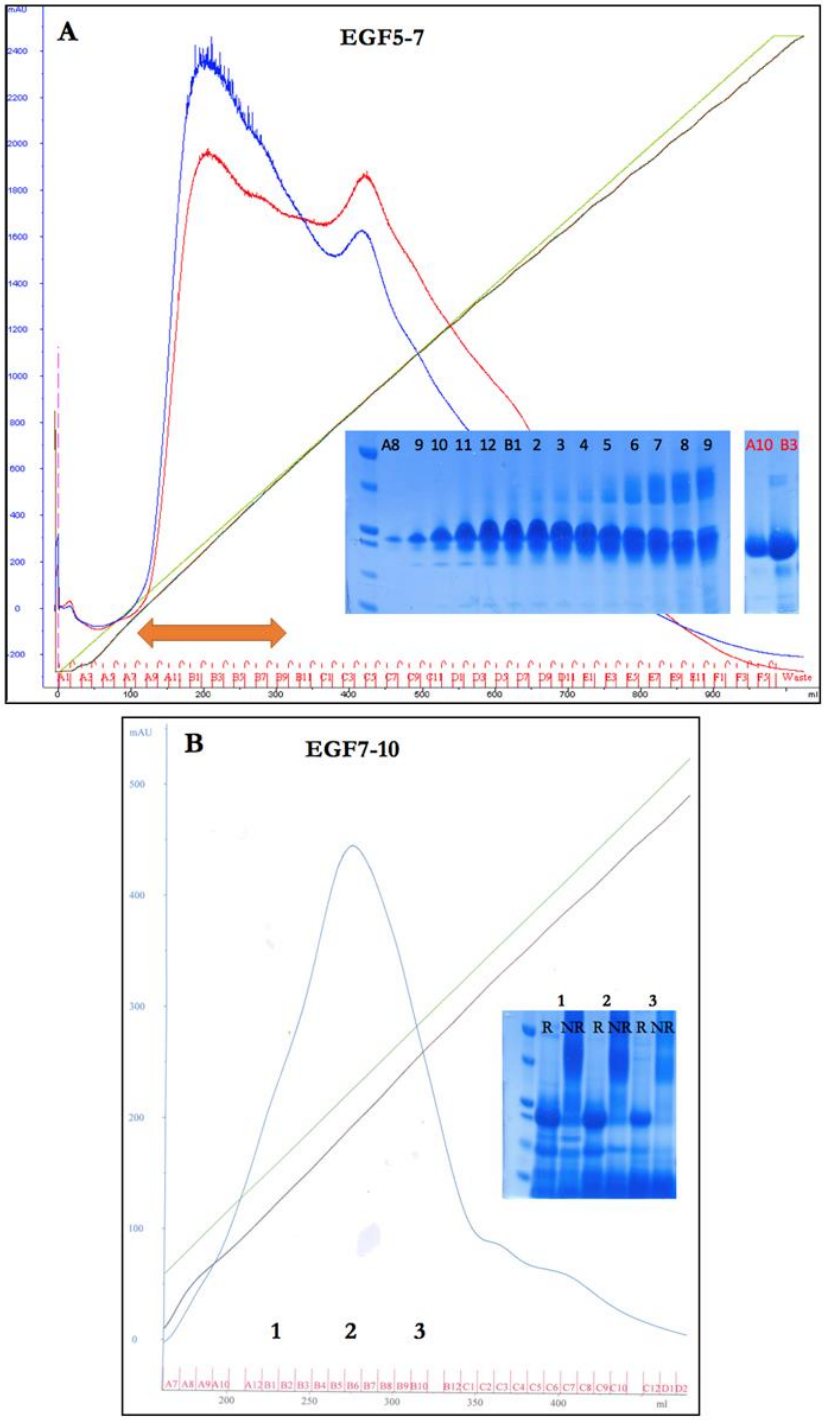


Figure 5.4. Cation Exchange Chromatography of EGFs 5-7 and 7-10: A. Chromatogram of EGF5-7 with 5 µl of the key fractions taken from the region of the orange arrow analysed via SDS-PAGE (black labelled are under NR conditions and red is with RSB). B. Chromatogram of EGF7-10 with 25 µl of the key fractions analysed via SDS-PAGE under R and NR conditions. Fraction 1 corresponds to B1, 2 to B6 and 3 to B10.

After concentration, the samples underwent further purification via SEC (data not shown). The total yield of the EGF5-7 monomeric fractions was 22 mg, and for EGF7-10 the total yield was 27.6 mg of protein obtained in total.

Due to a broken 3 L fermenter, and the considerable expense of 10 L of minimal media isotopic reagents, a shaker flask method was utilized to produce 250 ml of ^{15}N-labelled EGF5-7 and 7-10. A variety of conditions were tried in attempts to optimize the yield of protein, including incubator shaker speed (180 versus 220 rpm), the type of ^{15}N ammonium used (chloride versus sulfate) and induction at 18 °C instead of room temperature (data not shown). The best results were from using the conditions underlined above (figure 5.5). ^{15}N-EGF5-7 could be produced at around 0.35 mg/ml of supernatant (approximately a third of that which was produced unlabelled), whilst there was a more significant decrease in yield for the ^{15}N-EGF7-10 (in contrast to unlabelled protein) of around 0.1 mg/ml; a ten-fold decrease. However, in a volume of 250 ml, this meant that there would still be enough protein for NMR-based analysis (with an estimated 8 mg in 1 ml and 2.5 mg in 1 ml for ^{15}N-EGF5-7 and ^{15}N-EGF7-10, respectively).

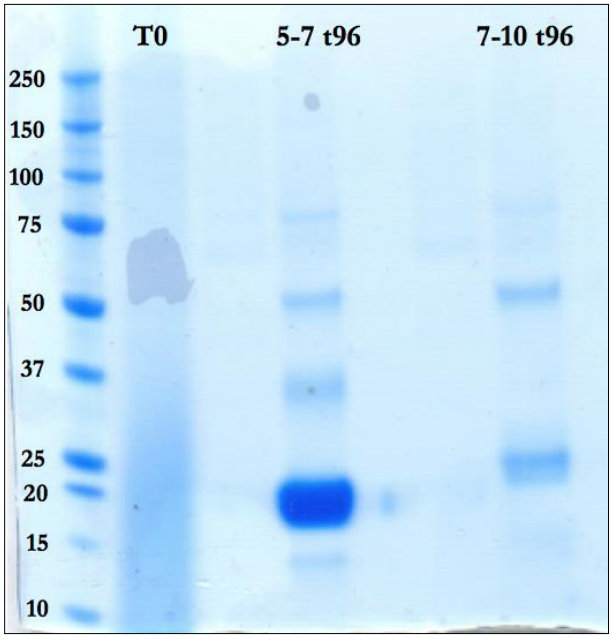


Figure 5.5. SDS-PAGE of ^{15}N-labeled EGFs 5-7 and 7-10: 30 µl of each sample (after 96-hours post induction) was loaded onto each lane. T0- the NI *Pichia* sample. All samples were run under reducing conditions.

The supernatant from each protein production run was purified via cation-exchange chromatography using SDS-PAGE to analyse the key fractions (data not shown). Once the relevant fractions of ^{15}N-EGF5-7 and ^{15}N-EGF7-10 were combined and concentrated, SEC was carried out for further sample purification (required due to contaminating *Pichia* proteins). This allowed for the majority of the proteins of interest to be isolated (data not shown); however, the presence of minor proteolytic fragments of only a few amino acids difference of ^{15}N-EGF5-7 warranted a final purification step that was to be carried out via RP-HPLC.

The ^{15}N-EGF5-7 was applied to RP-HPLC (several gradients were used for testing, ultimately settling on a shallow gradient), giving rise to an unusual elution profile on the chromatograph (figure 5.6). However, significantly improved separation of monomeric protein was achieved, with fractions F11 – F3 undergoing lyophilization and resuspension in 20 mM sodium acetate pH 4, 50 mM NaCl. Two protein concentrations were tested via SDS-PAGE to ensure the

final product was not only devoid of contaminating proteins, degradation products or dimers, but to also assess if the protein would precipitate at a high concentration (figure 5.7).

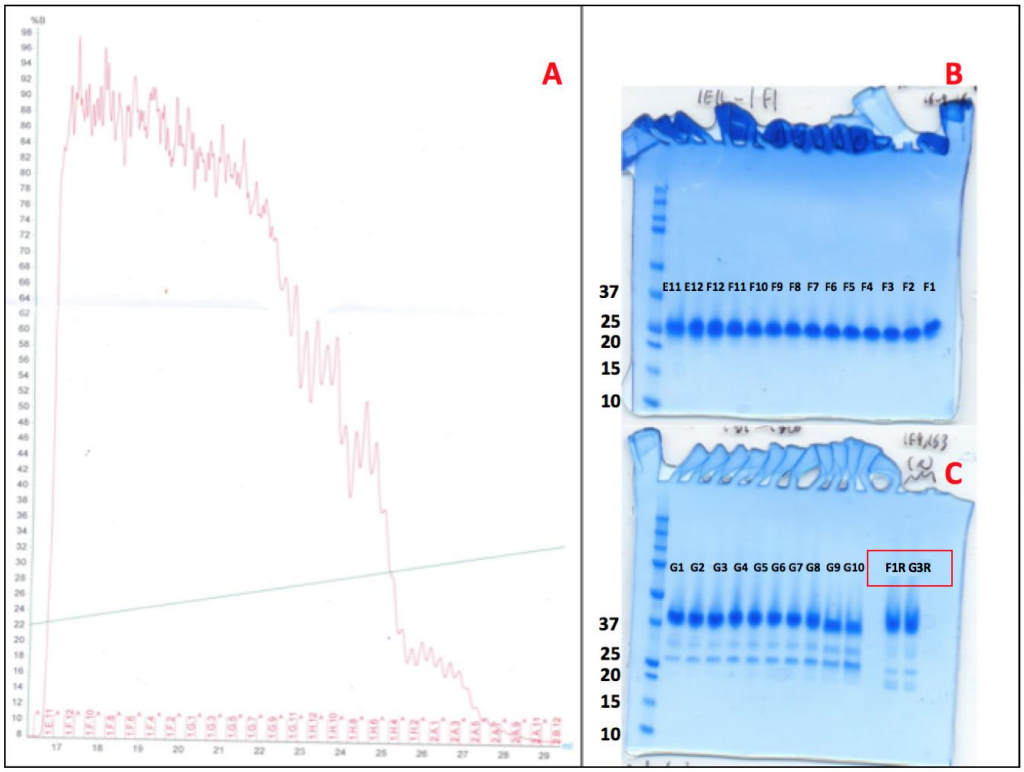


Figure 5.6. RP-HPLC of ^{15}N-EGF5-7: A. Chromatograph showing elution of the protein begins around 16% ACN. B. NR samples (3 µg) of fractions E11 – F1. C. NR samples (3 µg) of fractions G1 –G10 and R samples for fractions F9 and G3 (red box). The mAU reading of the highest peak point was ~1500.

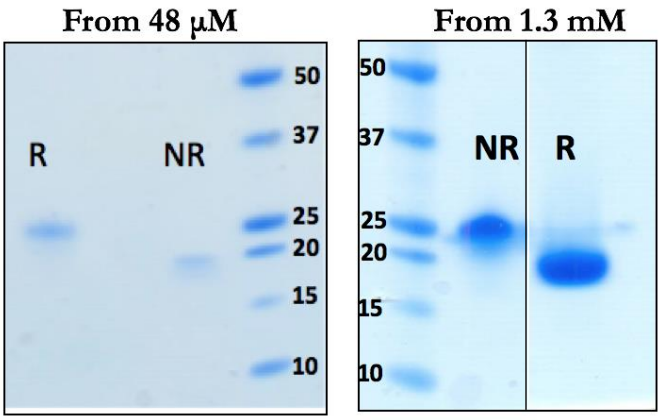


Figure 5.7. SDS-PAGE Analysis of Final ^{15}N-EGF5-7 Sample: 0.5 µg of protein was assessed under both R and NR conditions from a 48 µM sample of ^{15}N-EGF5-7. 3 µg of protein was assessed under both R and NR conditions from a 1.3 mM sample of ^{15}N-EGF5-7 (this gel is a composite).

^{15}N-EGF7-10 was split into two samples, one for further purification via cation exchange, and one for RP-HPLC (data not shown). Figure 5.7 shows the results of the RP-HPLC of ^{15}N-EGF7-10, showing pure, monomeric protein to be obtained. The key fractions from both chromatography runs were saved for future analysis by techniques such as NMR.

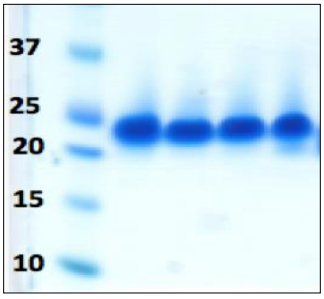


Figure 5.8. SDS-PAGE Analysis of Final ^{15}N-EGF7-10 Samples: 3 µg of protein was loaded under NR conditions. The samples were obtained from RP-HPLC purification of ^{15}N-EGF7-10.

Due to time and cost limitations, it was decided to pursue only EGF5-7 for structural analysis via NMR, therefore a double-labelled (^{15}N/^{13}C) sample had to be prepared. Before this, it was vital to ensure the final quality protein level that could be obtained in 250 ml of dll conditions (prepared in shaker flasks due to a

broken small-scale fermenter), that would yield a ^{15}N-EGF5-7 sample. The time points of the SS *Pichia* production of protein over three days is shown in figure 5.9. Based on the 20 µl volume of sample loaded, there was an estimated 65 mg of protein in the supernatant.

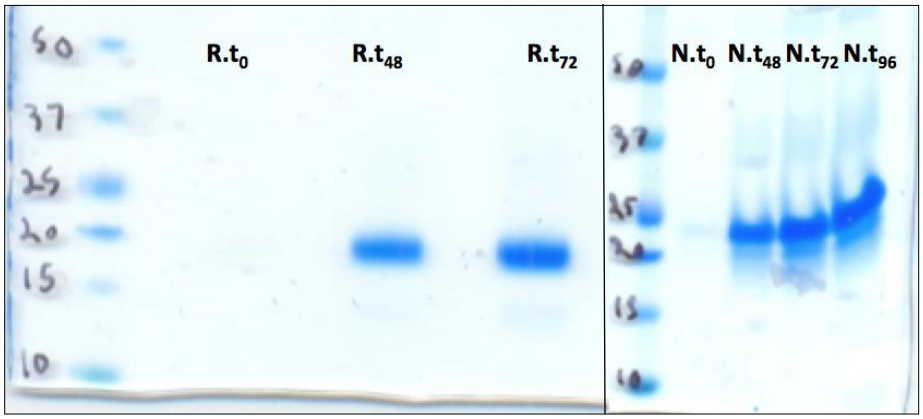


Figure 5.9. Induction of *SS Pichia* to Produce DLL ^{15}N-EGF5-7: 20 µl was loaded into each well. R refers to reducing conditions and N here refers to non-reducing. Time-points are given in hours (t with a subscripted number). The actual harvest occurred at 72-hours, with a small batch left to look at the 96-hour time point.

Following the harvest and dilution of the supernatant, cation exchange chromatography and subsequent SEC were carried out (data not shown). The cation exchange resulted in 31.5 mg of protein from combining and concentrating fractions C9 – D6 to 9 ml. Following SDS-PAGE analysis of the SEC fractions (figure 5.10 B), all the fractions were seen to contain some putative proteolytic products appearing as bands under the main monomeric protein. This was likely due to the use of less protease inhibitors during the production of the protein (only Pefabloc was added after 72-hours). Less inhibitors had been used since the actual double-labelled sample would be produced without them so as to not risk any interference with the uniform isotopic labelling of the recombinant protein.

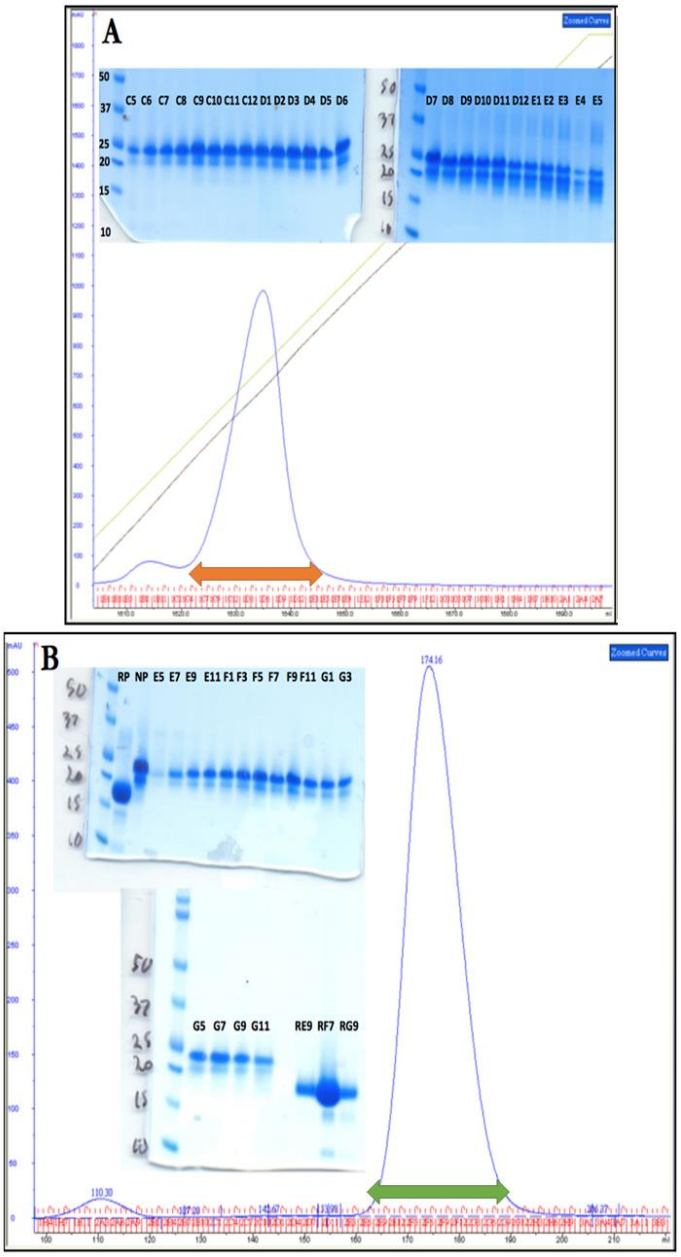


Figure 5.10. Purification of DLL-^{15}N-EGF5-7: A. 5 ml cation exchange column collecting 2 ml fractions from the area covering the orange arrow resulted in the chromatogram shown. The centre of the peak (fraction D7) contained an estimated 7.62 mg of protein. SDS-PAGE gels show the result of running ~3 µg of protein from each relevant fraction under NR conditions. B. SEC resulted in two peaks with the prominent eluting at ~174 ml (central peak fraction contains around 2.96 mg of protein; fractions taken from area covered by the green arrow). 2 µg of each protein fraction analysed via SDS-PAGE. Abbreviations: RP- reduced pre-SEC sample, NP- non-reduced pre-SEC sample. SEC column was an S75 26/600 preparative column collecting 1 ml fractions.

RP-HPLC was once again used to obtain ultra-pure monomeric protein that could withstand lyophilization and be reconstituted at a high concentration of > 1 mM (figure 5.11). Based on the protein quality, concentration obtained and the NMR analysis (data to be shown in section 4.4), it was decided to progress production of double-labelled EGF5-7

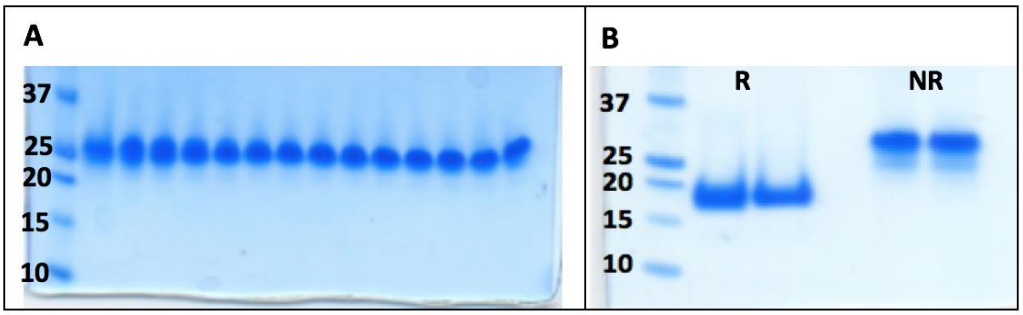


Figure 5.11. RP-HPLC Fractions and Final Sample DLL-^{15}N-EGF5-7 SDS-PAGE: A. Approximately 2 µg of protein from each relevant RP-HPLC fractions (chromatogram not shown) were loaded in with 1 x NRSB. These fractions were lyophilized and combined after being dissolved in 125 µl of an acetate buffer (20 mg/ml concentration). B. An estimated 3 µg of protein from the final reconstituted sample was run under both R and NR conditions. The first lane in each case contains an aliquot from the 20 mg/ml sample and the second lane is from a diluted sample of 5 mg/ml.

Copying the same production strategy used for the DLL sample, a cell culture was set up to produce ^{15}N/^{13}C-EGF5-7 and closely monitored (figure 5.12). The decision was made to harvest at 72-hours, following analysis via SDS-PAGE, because a fourth day was judged to unneeded and would likely just introduce protease-related issues (given that the culture remained free of protease inhibitor until after harvest). The protein concentration matched that seen in figure 5.9 (of the dll sample).

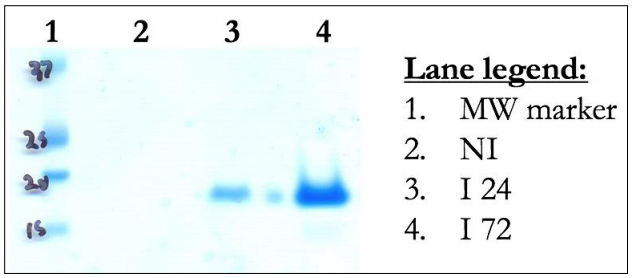


Figure 5.12. Induction of *SS Pichia* to Produce $^{15}N/^{13}C$-EGF5-7: 20 µl was loaded into each well. Abbreviations: NI- non-induced, I- induced (including the hours after induction).

Post-harvest, the sample was purified via cation exchange, then two rounds of SEC (data not shown). This procedure yielded mostly monomeric protein; however, the presence of minor proteolytic fragments warranted the sample undergo a final round of cation exchange purification. Relevant SEC fractions were combined, diluted 1:10 and applied to a 1 ml MonoS 5/50 GL column to attempt a high-resolution separation of the major and minor forms of the protein.

The results show that excellent separation was achieved, with ultra-pure monomeric protein being obtained, with separation of both minor fractions and those containing contaminating lower molecular weight products (figure 5.13). Fractions B7 – C1 were combined, buffer exchanged (20 mM sodium acetate, 50 mM NaCl, 0.01% sodium azide and 10% D_2O) and concentrated to a final volume of 350 µl at a concentration of 0.5 mM (a faulty batch of concentrators caused significant sample loss).

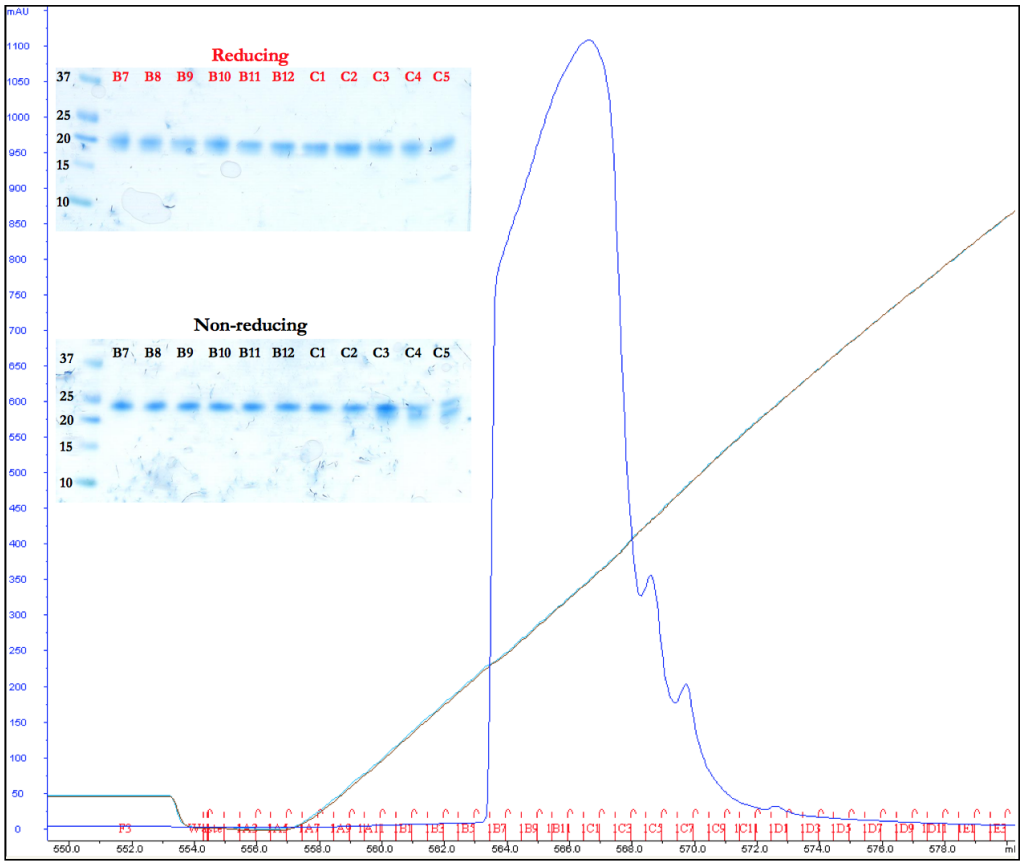


Figure 5.13. Final Purification of $^{15}N/^{13}C$-EGF5-7: The protein began eluting at around 15% buffer B (~150 mM NaCl) into 0.5 ml fractions. ~1 µg of each key fraction was analysed via SDS-PAGE under R and NR conditions (see inlaid figure).

After concentration, the protein was assessed via SDS-PAGE with a relatively high amount of protein loaded onto the gel (5 µg) under R and NR conditions so that the presence of significant break-down products, contaminants or dimer could be assessed (figure 5.14). The protein demonstrated none of these issues and appeared to be highly pure monomer. This sample of $^{15}N/^{13}C$-EGF5-7 was stored at 4 °C until NMR analysis was carried out several hours later.

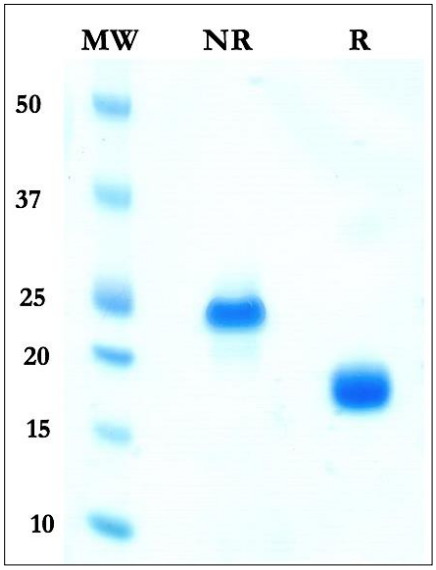


Figure 5.14. Final $^{15}N/^{13}C$-EGF5-7 Sample: ~5 μg of protein was used to load onto the gel under R and NR conditions.

5.4 Biophysical Characterization of EGF-Rich Modules

5.4.1 Methods for the Biophysical Characterization of EGF-Rich Modules

5.4.1.1 Mass Spectrometry

Mass spectrometry was carried out as per section 3.4.1.1 at the SIRCAMS facility and St. Andrews University.

5.4.1.2 Circular Dichroism

CD was carried out as per section 3.4.1.2 at the University of Edinburgh with assistance from Dr. Elizabeth Blackburn. The CD Analysis & Plotting Tool (CAPITO) from Leibniz-FLI (http://capito.nmr.leibniz-fli.de/) was used for data

plotting for data with buffer subtracted. Dichroweb software (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) used for data analysis.

5.4.1.3 Thermal Analysis via TSA and NMR

Thermal stability was assessed, as per section 3.4.1.3, at the University of Edinburgh using either a BioRad IQ5 ICycler (with experimental help provided by Dr. Elizabeth Blackburn) or by NMR (with the assistance of Juraj Bella for experimental set up).

5.4.1.4 1D and 2D NMR Analysis

NMR was carried out as per section 3.4.1.5 at the University of Edinburgh with experimental set up performed by Juraj Bella, and advice given by Profs. Paul N Barlow and Dusan Uhrin, and A/Prof. Matthew Call.

5.4.1.5 Crystallography

Crystallography was carried out, as per section 3.4.1.6, at the EPPF. Ella Svahn aided the initial crystallization trials with advice provided by Dr. Atlanta Cook. Six 96-well plates were screened for crystal growth under various conditions (provided by the EPPF; Index, Structure Screen, JCSG+, MIDAS, PEGION and Salt Rx) using a highly purified sample of EGF5-7 at a concentration of 15 mg/ml in a buffer consisting of 20 mM sodium acetate, pH 4, and 50 mM NaCl. Crystallisation trials were also carried out on a sample of 10 mg/ml EGF7-10 in the same buffer. A follow up trial was performed on a 25 mg/ml sample of EGF5-7.

5.4.1.6 Homology-Based Modelling and SAXS Methods

Homology modelling was carried out using the I-TASSER Protein Structure & Function Predictions software suite (https://zhanglab.ccmb.med.umich.edu/I-TASSER/), and Swiss-Model software developed for the Biozentrum Universitat Basel Center for Molecular Life Sciences (https://swissmodel.expasy.org/interactive). Proteins sequences were edited to have cloning artefacts removed. Advice on modelling was provided by Dr. Dinesh Soares.

A 0.5 mM sample of EGF5-7 in 20 mM sodium acetate pH 4 and 75 mM NaCl was prepared and analysed by Dr. Haydyn Mertens of the BIOSAXS group at the European Molecular Biology Laboratory (EMBL), Hamburg (Germany). The methods utilized a 1M PILATUS pixel detector (Dectris, Switzerland) and eight frames of 15-s exposure time. The sample-to-detector distance was 2.7 m, covering a range of momentum transfer $0.08 \leqslant s \geqslant 6.0$ nm-1 (where $s=4\pi\sin(\theta)/\lambda$ with 2θ being the scattering angle and $\lambda=0.15$ nm being the wavelength). Dr. Mertens performed the processing of the data, with analysis carried out with the assistance of the author (fit of the homology model to the SAXS data, analysis of flexibility, and *ab initio* shape determination).

5.4.1.7 Assessment of Free Cysteines and Disulfide Mapping

Samples of EGF5-7 were provided to Dr. David Clark and Lavrentis Galanopolous for a mass spectrometry-based analysis of the presence of free cysteines and for disulfide mapping. Briefly, N-ethylmaleimide (NEM, Thermo Fisher) was used to mix with EGF5-7 (either with or without tris[2-carboxyethyl]phosphine, TCEP) prior to ESI-MS analysis (methods as per (Hill, *et al.*, 2009). Disulfide mapping was attempted with trypsin and pepsin according to

the general methods outlined in (Hodder, *et al.*, 1996). All reagents were purchased from Sigma-Aldrich unless otherwise stated. All solvents were LC-MS grade unless stated otherwise. More specific methodology below was provided kindly by Lavrentis Galanopolous.

The EGF5-7 protein (500 µl, 1 mg/ml) was pre-concentrated by ultrafiltration. The preconcentrated solution (33 µl, 3 mg/mL) was added to a PBS pH 6.8 solution (77 µl, 10 mM NaCl). A 50 µl aliquot was added to ammonium bicarbonate solution (NH_4HCO_3) (50 µl, 50 mM). Trypsin (1 µg/µl) was added to give an enzyme to protein ratio of 1:50 and 1:10 by weight. The resulting solution was incubated at 37 °C overnight. For the peptic digest, the pre-concentrated protein solution was diluted into a PBS pH 1.5 solution (50 µl, 10 mM NaCl). Pepsin (1 µg/µl) was added to give an enzyme to protein ratio of 1:50 and 1:10 by weight and the solution was incubated at room temperature overnight.

Alkylation reagent stock NEM and TCA were stored as 100 mM and 20 % (w/v) solution (in water), respectively, at -20 °C. For protein alkylation with NEM, the stock solution was diluted to either 5 or 20 mM concentrations, and the alkylations were conducted for half an hour at room temperature in the dark. For protein alkylation with TCA, the TCA stock was diluted in a 1:10 and 1:2 ratio to protein solutions and the alkylations were conducted for 15 min at 75 °C in the dark.

ZipTip® (EMD Millipore) pipette tips were used before the injection of the proteolysed samples in the FT ICR MS. This ensured sample desalting, concentration and purification. Each C_{18} ZipTip was pre-washed twice with 100% ACN (100 µl) and 0.1 % formic acid (100 µl). The sample was passed through the tip ten times and then washing twice with 0.1 % formic acid (100 µl). Finally, it was eluted in 50 µl of 66 % ACN 0.1 % formic acid.

All LC/MS-ESI carried out in the SIRCAMS facility as described in section 3.4.1.1, using a solariX FT-ICR mass spectrometer coupled to a 12 Tesla superconducting magnet (Bruker Daltonics) and Ion Mobility Q-TOF Waters Synapt spectrometer (Bruker Daltonics). The apparatus was run in the positive-ion mode with a capillary voltage of 4 kV. The drying gas was set to 180 °C and 4-L/min. The nebulizer gas pressure was 5 bar. Parameters were optimised according to the experiment under investigation using the operating software, solariX control. FT-ICR data were interpreted using Data Analysis (Bruker Daltonics). MS/MS was also performed on a solariX FT-ICR mass spectrometer. Specific ion species were isolated employing the mass resolving quadrupole (Q), and MS/MS was conducted applying either collision-induced dissociation (CID) or electron capture dissociation (ECD). Ion accumulation time was raised to at least one- to two-seconds to escalate ion abundance. Fragmentation spectra were the sum of 50 - 100 acquisitions over an m/z range of 200 - 3800.

5.4.2 Results of the Biophysical Characterization of EGF-Rich Modules

5.4.2.1 Mass Spectrometry

To ascertain whether the sample had the expected mass and to estimate the number of disulfide bonds present, a 30 µl aliquot of EGF5-7 was analysed via LC/ESI-MS (figure 5.15). With an error margin 0.1234 Da, the MW matched up perfectly with the simulated mass that was calculated based on subtracting 20 Da (for each assumed disulphide bond formed between Cys residues) from the predicted MW of the sequence in Expasy ProtParam and subtracting the Ala-Gly cloning artefact that *P. pastoris* sometimes adds to the N terminus of the protein due to incomplete Kex cleavage. The results suggest that all 20 Cys residues are forming disulfide bonds, including those found in the predicted long linker region between EGFs 5 and 6.

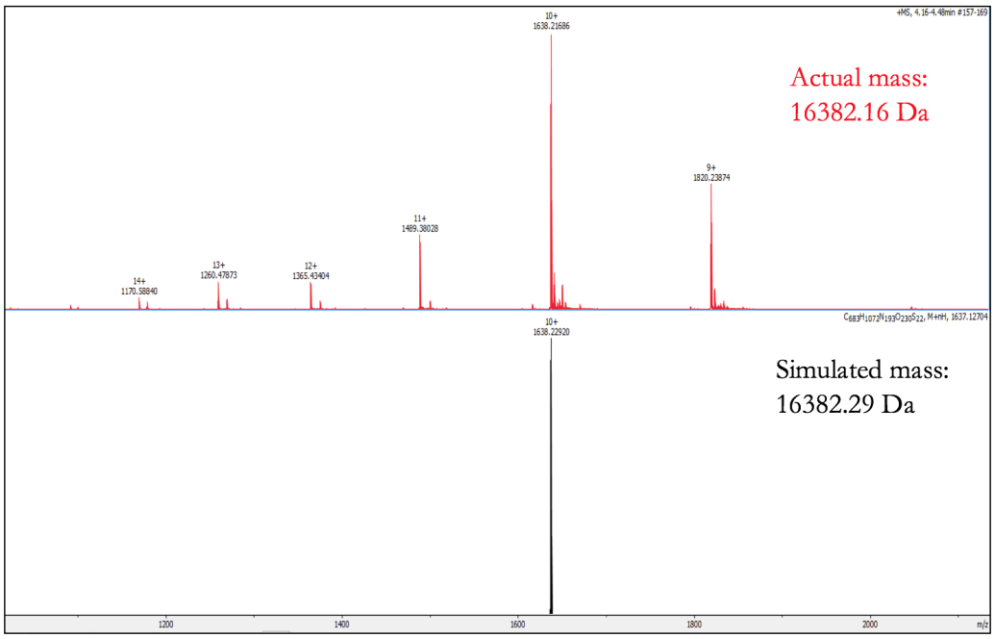


Figure 5.15. LC/ESI-MS Analysis of EGF5-7: Shown in red is the actual mass obtained in a series of charge states. Below, in black, is the simulated mass predicted based on all 20 cysteines forming disulfide bonds.

For correct mass identification of EGF7-10, MALDI-MS was used (due to machine availability). A 30 µl aliquot of EGF7-10 was utilized (figure 5.16). With an error margin of 0.11 Da, the MW matched up perfectly with the simulated mass that was calculated based on subtracting 24 Da from the predicted MW of the sequence in Expasy ProtParam and subtracting the Ala-Gly cloning artefact as before for EGF5-7. The results suggest that all 24 Cys residues are forming disulfide bonds, which is expected from a protein comprised of four EGF-like domains in tandem.

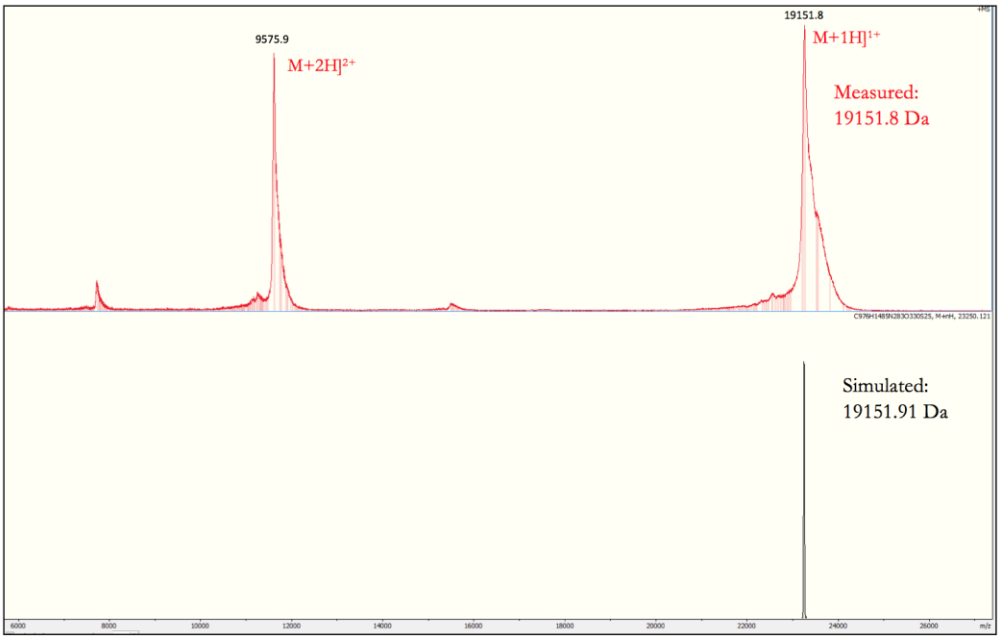


Figure 5.16. MALDI-MS of EGF7-10: Shown in red is the actual mass obtained in a series of charge states. Below, in black, is the simulated mass predicted based on all 24 cysteine residues forming disulfide bonds.

Once the ^{15}N-EGF5-7 sample had been produced, it was likewise important to ensure correct mass identification. However, due to the ^{15}N isotope uniformly labelling the protein, the mass estimation would be less accurate. In this case it was more important the mass matched overall (between 16.3 and 16.5 kDa, depending on the proportion of protein with cloning artefact present, and assuming an enrichment of 99.9%). Figure 5.17 shows the results of the LC/ESI-MS of a sample of ^{15}N-EGF5-7, demonstrating the roughly expected mass of 16.5 kDa, corresponding to the predicted mass of the protein where all nitrogen atoms possess an atomic mass of 15 instead of 14.

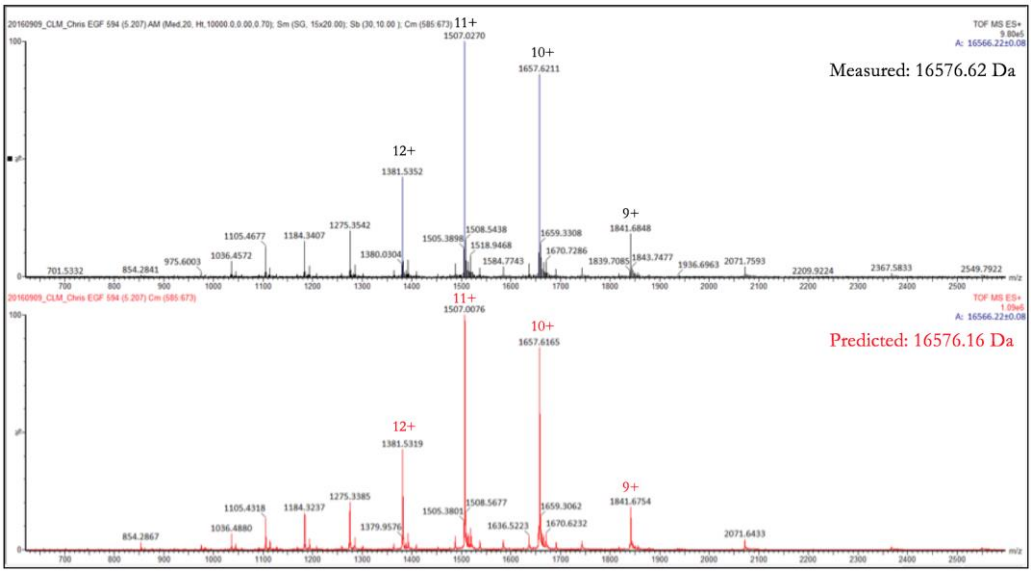


Figure. 5.17. LC/ESI-MS Analysis of ^{15}N-EGF5-7: Shown in black is the actual results and in red are the predicted results of the mass analysis. Key charge states are given above their peaks.

5.4.2.2 Circular Dichroism

CD was employed to assess the extent and nature of secondary structure in the *P. pastoris*-produced EGF5-7. The data could be compared with predictions on the basis of homology and with CD data for EGF5-6 produced in *E. coli*. Further, whilst EGF5-7 should be folded by way of expression and secretion from SS. *Pichia*, it is always good practice to ensure there is evidence for protein folding via CD and NMR (an activity assay should be additionally used if producing an enzyme or a protein with a known binding partner). The far-UV CD spectrum on a sample of EGF5-7 (figure 5.18 A) had a minimum at ~198 nm, and was similar to the CD spectrum collected on EGF5-6 (figure 5.18 B). As expected, analysis of the spectrum predicted that β-strand is the predominant secondary structure (over 35%), with less than 5% predicted to be α – helical. The remainder comprised of ~19% turns, and 41% disordered or unstructured (41%). However, the NRMSD value was several orders of magnitude higher than that seen for the CD of EGF5-6 (figure 4.22), which reflected a poor fit of the spectra obtained compared to the

reference spectra used by Dichroweb (reference set 7). For this reason, a set of reference spectra from CAPITO (a web server-based analysis and plotting tool used for CD data) were identified that had similar levels of α-helix, β-strand and unordered regions. This gave an improved NRMSD value of up to 0.92. Finally, the EGF5-7 protein overall takes on the same CD curve shape as the *P. falciparum* triple EGF repeat protein fragment, cbEGF7-9, as shown in figure 4.21 D, albeit with a slight positive maxima around 227 nm, possibly a contribution from a small α-helical region. Differences between these two proteins could also be due, in part, to the long linker found between the 5th and 6th EGF domains of EGF5-7.

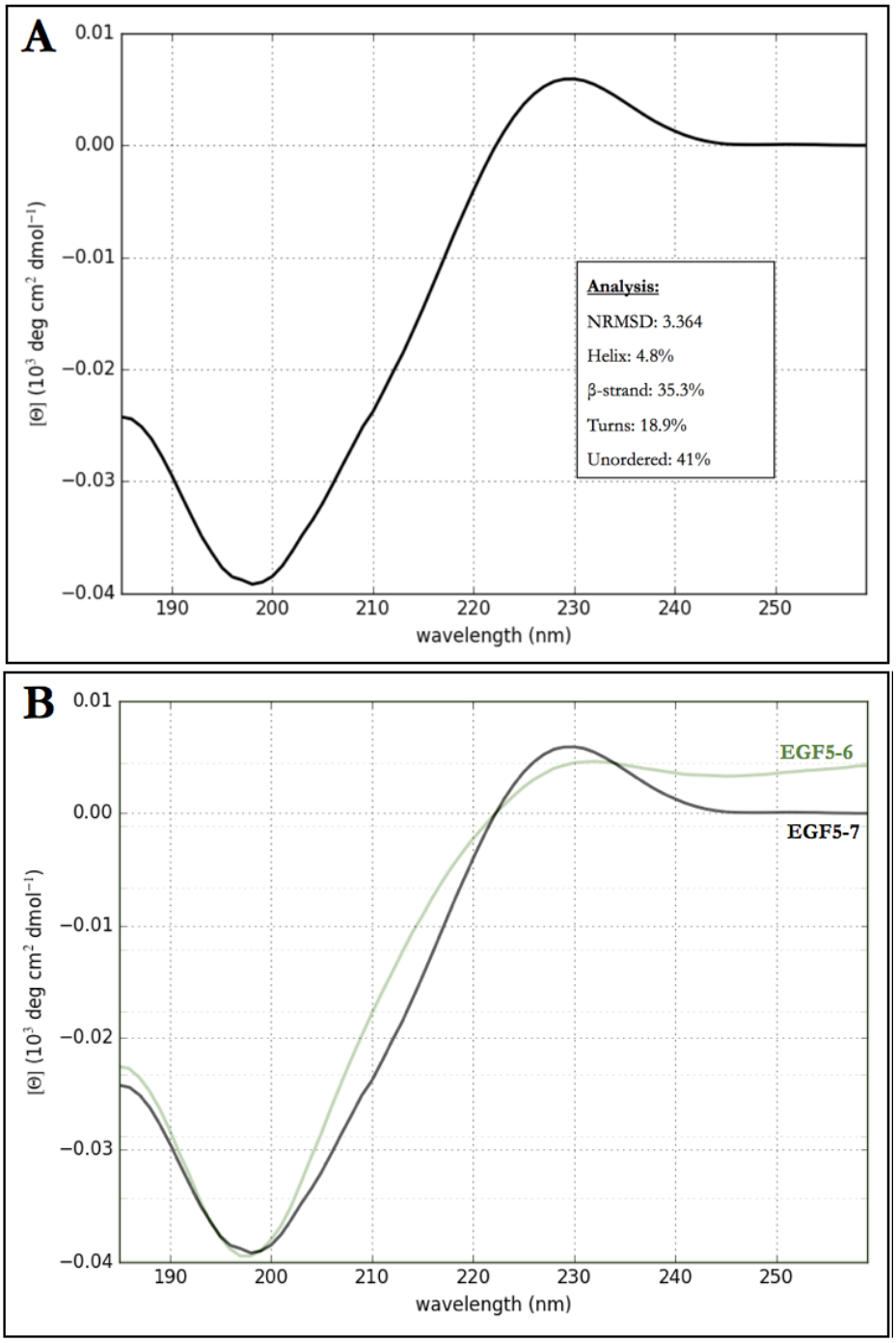


Figure 5.18. Far UV CD Spectra of ^{15}N-EGF5-7: A. The CAPITO-generated CD spectrum for EGF5-7 including quantitative analysis information from Dichroweb. B. Comparison of CD spectra from EGF5-7 (black) and EGF5-6 (green).

5.4.2.3 Thermal Analysis via TSA and NMR

Unlike *E. coli*-produced EGFs 3-6 and 5-6 (see figure 4.23), *P. pastoris*-produced EGF5-7 did not appear to denature in the TSA (data not shown). This suggested that the protein was either not binding the dye molecules to give a significant relative fluorescence signal, or the addition of EGF7 made the protein significantly more thermostable. Due to the possibility of the former, two tests were conducted; boiling an aliquot of EGF5-7 for one-minute to see if any precipitant could be observed (thus indicating the lack of thermostability of the protein), and assessment of the sample via NMR, with the temperature being increased every 30-minutes and a record of the HSQC obtained. With respect to the former, no precipitation was observed by eye.

With respect to NMR analysis (figure 5.19 – 5.21), the spectrometer was only able to go from 25 °C to 70 °C due to the need to avoid damage to the probe at higher temperatures, so an exact temperature comparison to the TSA was unable to be carried out. Figure 5.19 shows how each cross peak (corresponding to an NH group, mostly in the backbone) migrated across the spectrum over the course of the experiment, moving from left to right as the temperature increased. This reflects the change in referencing relative to the water, whose peak position is temperature sensitive, not a genuine change in chemical shift, despite movement occurring in both the ^{1}H and ^{15}N dimensions. As the temperature increased, peaks appeared much faster, with the most pronounced changes occurring at 70 °C. This is particularly seen when comparing and contrasting the HSQC data sets obtained at 40 °C and 70 °C. Both side chains and various residues had appeared from the spectra at the latter temperature, suggestive of changes in the local chemical environment, contributed in part from certain amide regions undergoing faster exchange. Alternatively, these changes are reflective of the sensitivity of the water peak to the temperature (likely changing more rapidly between 65 and 70 °C), and

so the overall evidence of melting up to 70 °C is lacking, thus indicative of a highly thermostable protein.

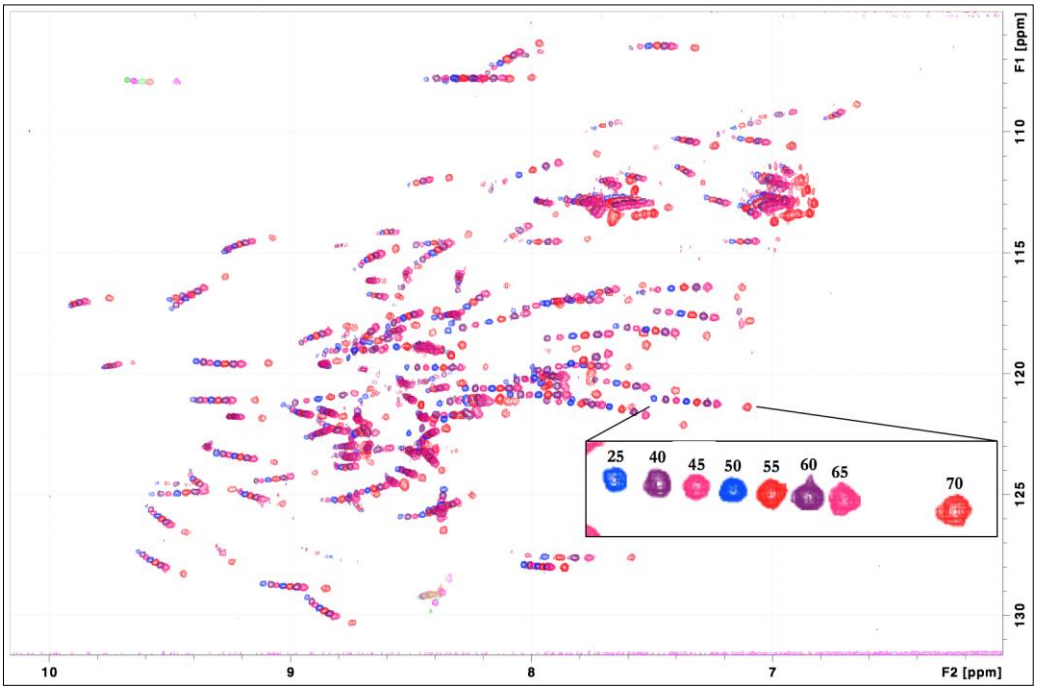


Figure 5.19. NMR-Based Temperature Shift for ^{15}N-EGF5-7: HSQC spectra were obtained from a ^{15}N-EGF5-7 sample over 30-minutes for each temperature (25 and 40 – 70 °C). An example of the peak position at each temperature for one backbone amide group is shown inlaid in the figure. These spectra include folded peaks.

Despite the excellent quality of HSQC spectra at 70 °C, consistent with fully folded protein, there was evidence of a cloudy precipitate in the sample tube after exposure to this temperature. Examination of 1D spectra (figure 5.20) likewise indicated potential evidence for temperature-related loss of folding. Focusing on the amide region of the 1D NMR, the relative intensity of the peaks from 6.25 – 9.5 ppm diminishes as the temperature increases, with up to 75 % less peak intensity at 70 °C compared to the spectra obtained at 40 °C (a specific example would be the peak at 8.5 ppm). In the methyl region, a complete abolition of the peak at ~0.25 ppm occurred by 60 °C. However, no significant peak intensity alteration is discernible in the β-sheet backbone Hα region around 5 – 5.5 ppm

(except for a peak at ~5.25 ppm in spectra collected at 65 – 70 °C). This is somewhat expected given the EGF domains are comprised of β-sheet with disulfide bonds that are generally high temperature resistant. This is indicative and consistent with a proportion of protein no longer in 100% native state that is precipitating out of solution (or at least a minor proportion of it doing so).

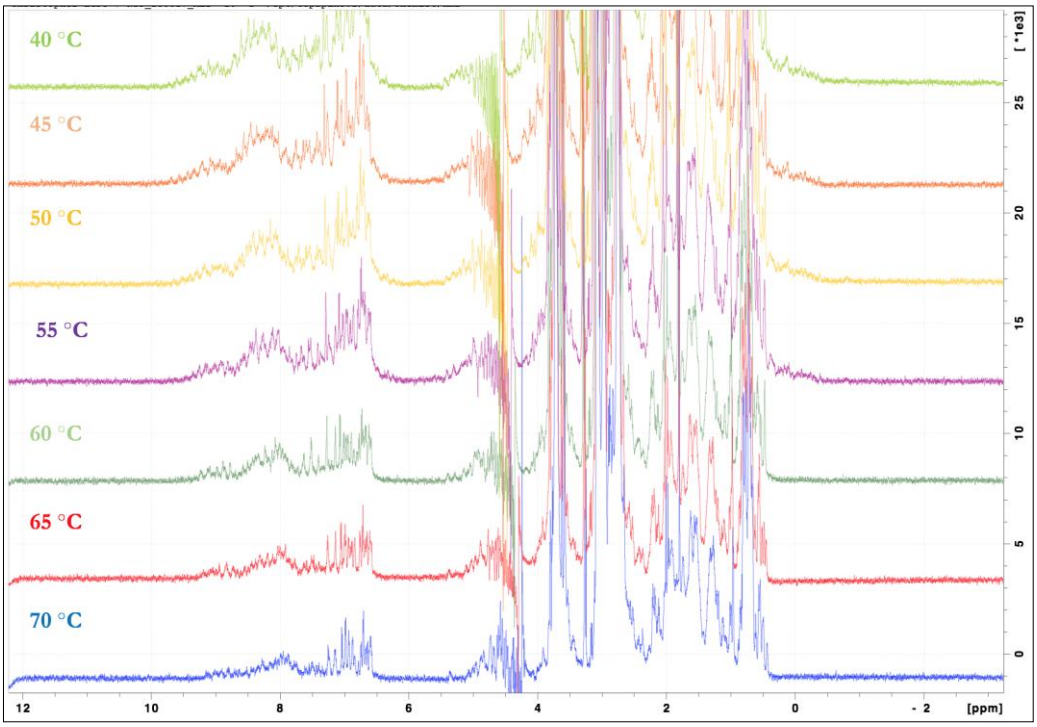


Figure 5.20. 1D NMR Analysis of ^{15}N-EGF5-7 from 40 to 70 °C: Each spectrum is matched in colour to the temperature at which the experiment was performed.

5.4.2.4 1D and 2D NMR Analysis

Due to the aim of pursuing structural analysis via NMR, it was important to assess how EGF5-7 behaved under a variety of conditions such as pH/buffer type, salt concentration, protein concentration and temperature. First, a variety of buffers were tested to assess the impact of pH on the sample. This is important since a low pH generally results in better quality NMR spectra as it increases the rate of

backbone amide proton exchange with solvent. Despite this, a pH as high as 6.5 or 7.5 is desirable as it emulates physiological conditions. In figure 5.21, the sharpness and resolution of peaks at the lower pH (as expected). It appears that acetate and MES buffers give better spectra than PBS.

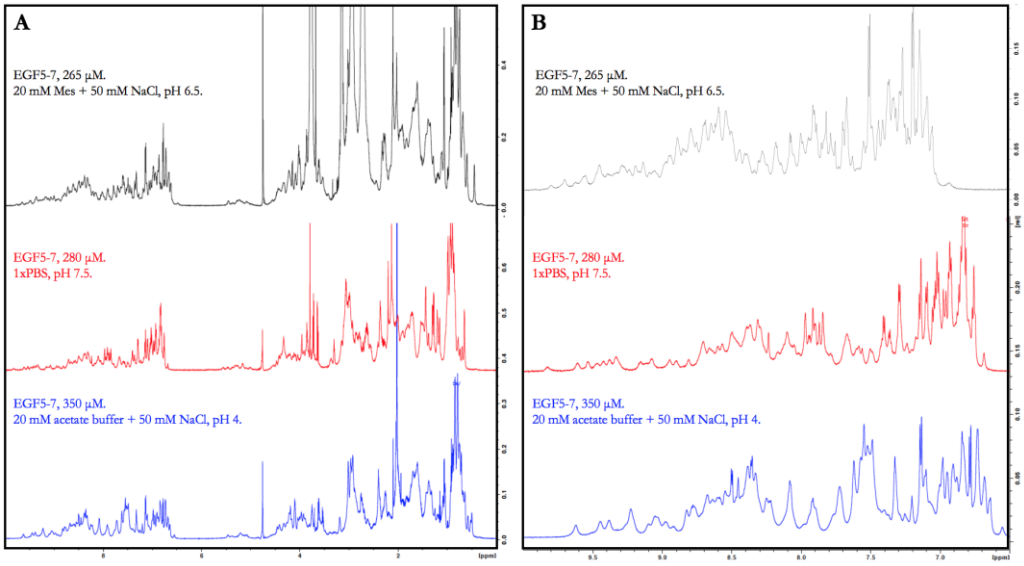


Figure 5.21. Effects of pH/Buffer on ^{15}N-EGF5-7: A. 1D NMR spectra of EGF5-7 in different buffers with different pH levels, as indicated by the colour coding. B. Zoom-in on the amide region.

Screening to ascertain the minimum ionic strength of the buffer that our proteins can tolerate can be important for obtaining optimal signal-to-noise, especially when using a cryoprobe, which is adversely affected by salt. When a sample of EGF5-7 was dialysed into a zero-salt buffer, a precipitant quickly formed. Hence, 50 mM and 100 mM NMR buffers were utilised. As shown in figure 5.22, 50 mM NaCl sample at pH 4 produced a high-quality 1D ^{1}H spectrum. It was decided to use this concentration for future NMR experiments in order to increase the ease of tuning and matching of the solvent deuterium signal, and to reduce the longer 90° pulse durations with the use of the cryoprobe.

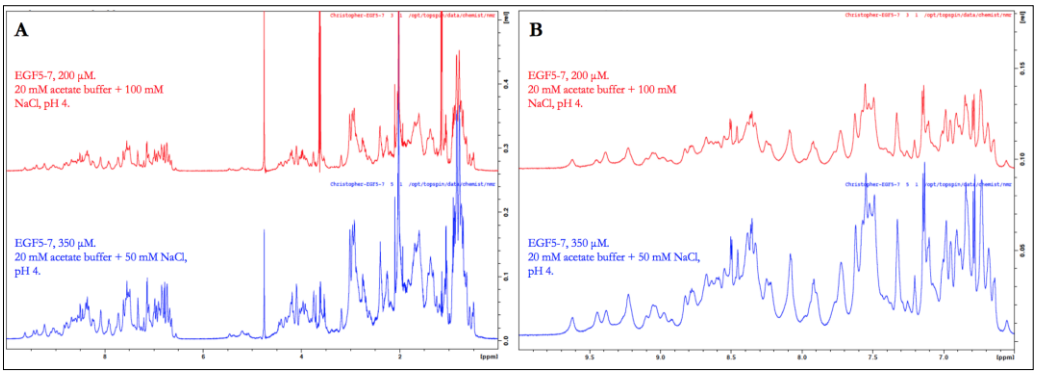


Figure 5.22. Effects of Salt Concentration on ^{15}N-EGF5-7: A. 1D NMR spectra of EGF5-7 in 100 mM NaCl (red) and 50 mM NaCl (blue). B. Zoom-in on the amide region.

Having identified the best buffer, the next step was to assess the effect of protein concentration on collection of NMR spectra. Obviously, a higher concentration would cut acquisition times, but protein aggregation is more likely at higher concentrations. Initially, a 1.3 mM sample of EGF5-7 was analysed to obtain "standard" 1D, NOESY and HSQC spectra, with the latter experimental data collected on a 1:2 diluted sample (figure 5.23). The 1D spectra showed good peak dispersion, similar to that of the 0.35 mM sample (in the 50 mM NaCl buffer) in the same buffer in figure 5.22 A. The NOESY spectrum shows clear evidence of long-range methyl protons in spatial proximity with aromatic protons, as well as αH-NH cross peaks; both these features are consistent with a folded protein; consistent with prior HSQC spectra. Minor differences were seen between the HSQC spectra of a 1.3 mM sample and a 0.65 mM sample of EGF5-7 (figure 5.23 C and D respectively. Whilst both display well-dispersed peaks at the same locations, loss of peaks began to occur when the sample was at 1.3 mM, suggesting that mild self-association had potentially commenced. This indicated that concentrations below 1.3 mM should be used for 3D-structural data acquisition once a double-labelled sample was made.

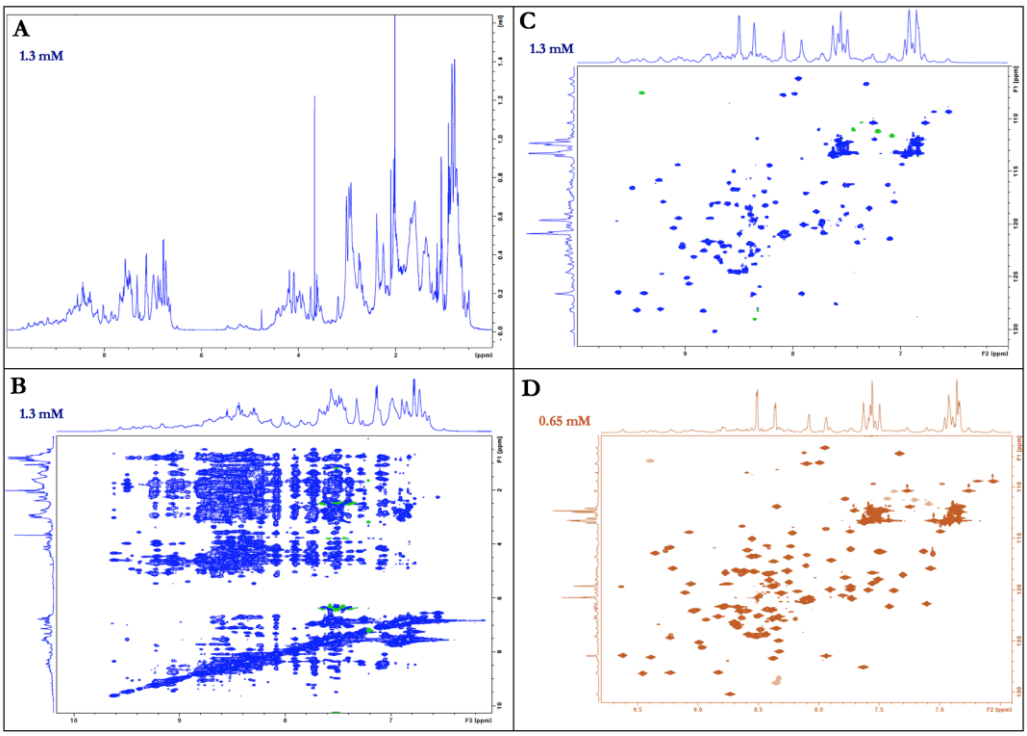


Figure 5.23. Effects of ^{15}N-EGF5-7 Concentration on NMR Data Obtained: A. .1D NMR spectra of EGF5-7 at 1.3 mM acquired over 0.65-seconds. B. 2D NOESY on the ^{1}H - ^{1}H plane of EGF5-7 at 1.3 mM. Mixing time was 120 milliseconds (msec), acquisition time was 0.1-seconds indirectly and 26 msec in ^{15}N plane plus 14.5 msec in indirect proton plane. C. 2D HSQC spectra of 1.3 mM EGF5-7 obtained over five-minutes. D. 2D HSQC spectra of 0.65 mM EGF5-7 obtained over 55-minutes.

Next, an assessment of temperatures and their impact on NMR data acquisition was required. Whilst previously (section 5.4.2.3), a range of higher temperatures (up to 70 °C) was explored to assess thermal stability, the goal here was to optimise, over a 25 to 37 °C range, the temperature to be used for data acquisition (figures 5.24A and 5.24B, respectively). At 37 °C, the peaks in the 1D spectra became significantly better resolved, particularly in the β-sheet backbone H α region (around 5 – 5.6 ppm). Likewise, the HSQC showed that the peaks sharpened up at 37 °C. However, spectra were judged to be almost as sharp and well-resolved at 25 °C. It was decided, on balance, to use the lower temperature (25 °C) for initial collection of 3D NMR data. This should lessen the danger of

proteolytic degradation that might occur over the extended period (up to a month) required to acquire the full suite of spectra needed for structure determination. Subsequently, if the sample remained intact, other temperatures could be used for the collection of further NMR spectra. This strategy could provide a range of data sets that might be useful, particularly if there were significant issues with intermediate exchange at a particular temperature for a particular residue. Overall, the data generated here reveal a range of conditions that looked promising for acquiring spectra for a double-labelled sample of EGF5-7, particularly the variety of temperatures and protein concentrations that could be used.

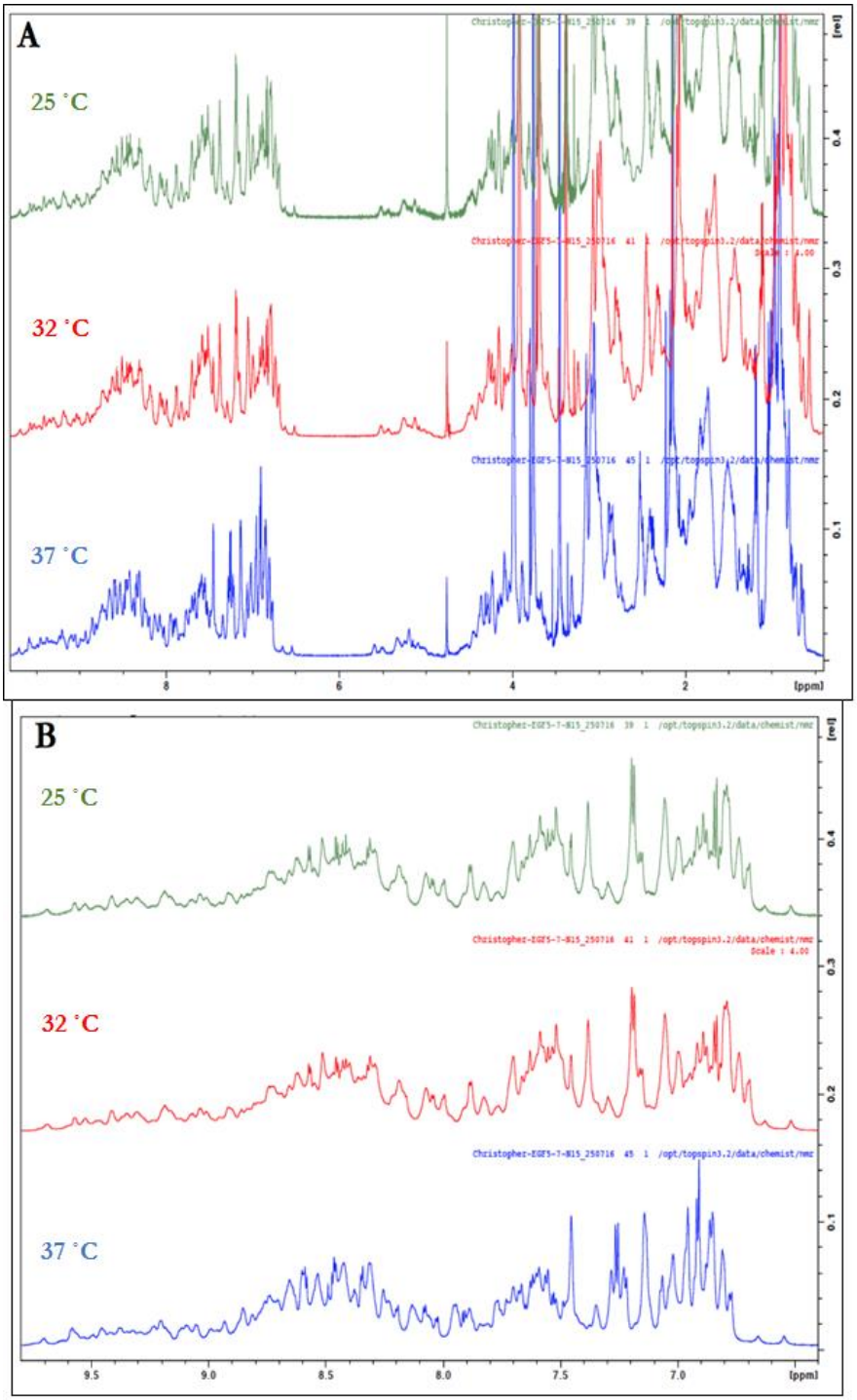


Figure 5.24A. 1D NMR of ^{15}N-EGF5-7 at 25, 32 and 37 ºC: A. 1D NMR spectra of EGF5-7 at 25 °C (green), 32 °C (red) and 37 °C (blue). B. Zoom-in on the amide region for each. 128 scans were taken of the 380 µM sample in 20 mM sodium acetate and 50 mM NaCl pH 4.

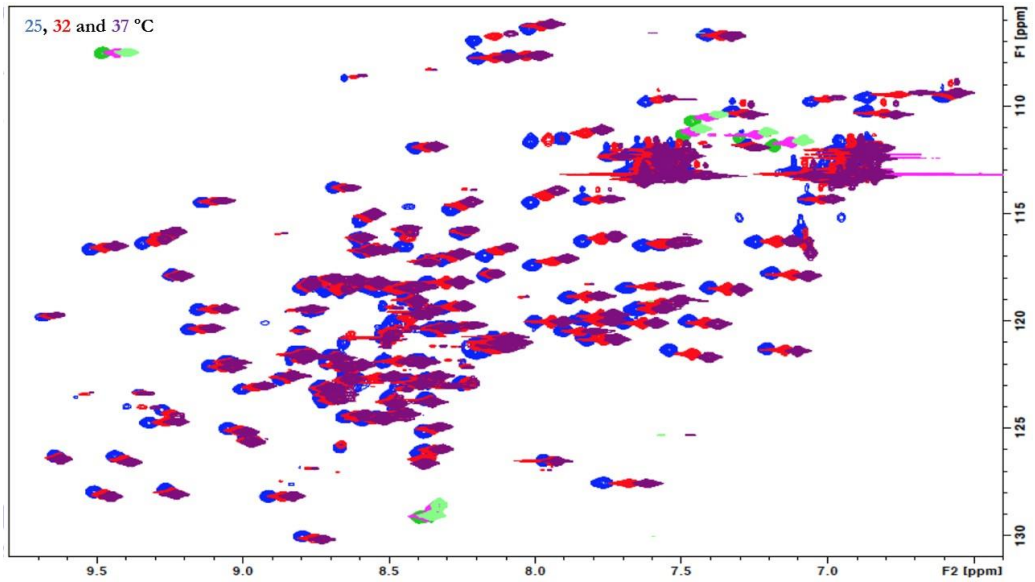


Figure 5.24B. HSQC of ^{15}N-EGF5-7 at 25, 32 and 37 °C: The HSQC shows the peak shift as temperature increases from 25 °C (blue) to 32 °C (red) and 37 °C (purple). The HSQC includes folded peaks (dark green, violet and bright green for temperatures 25, 32 and 37 °C, respectively). It should be noted that the spectra were referenced to water.

With respect to EGF7-10, a pH/buffer screen was carried out using 50 mM NaCl in each buffer (figure 5.25). The buffer with the lowest pH (the acetate buffer) yielded the best peak separation in the β-sheet backbone H α region (around 5.5 – 5.75 ppm). However, each EGF7-10 sample provided relatively poor-quality spectra when assessed in HSQC (for a temperature shift experiment like in figure 5.24, which also showed the peaks getting sharper at higher temperatures; data not shown) and particularly poor quality spectra from the NOESY experiments (data not shown). This was likely due to the presence of some minor aggregates and proteolytic fragments, in addition to the larger size of the protein (~20 kDa, which may be elongated thus having a reduced tumbling speed). Due to this, and the time constraints, it was decided to focus on EGF5-7 for further work, including preparation of a double-labelled sample.

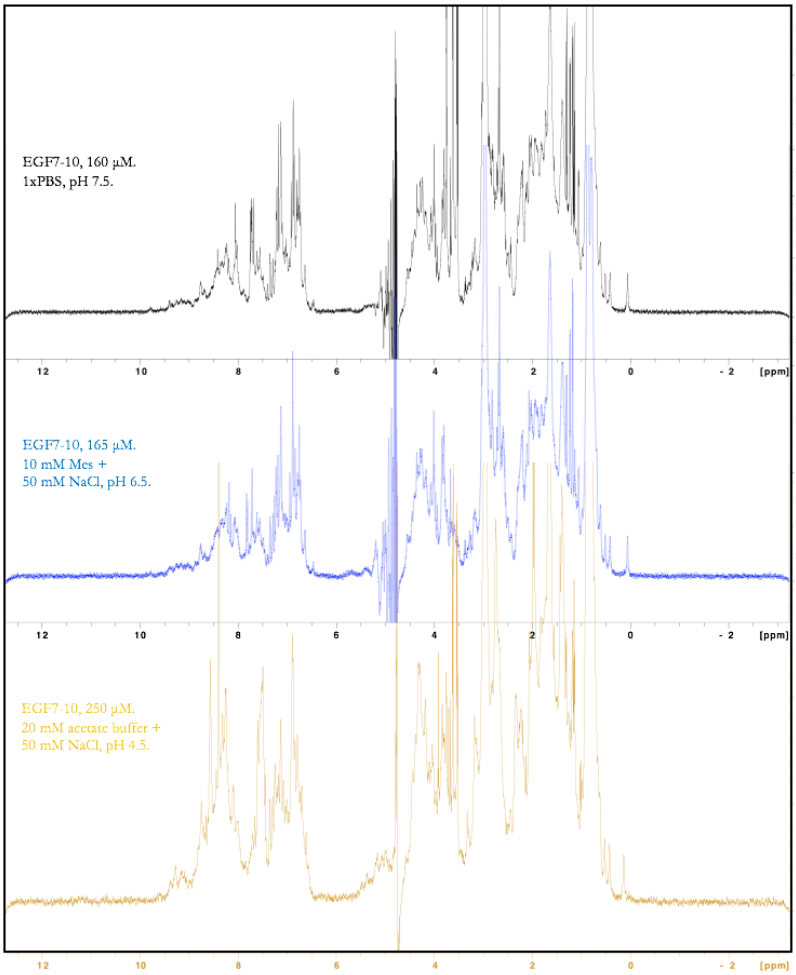


Figure 5.25. Effects of pH/Buffer on ^{15}N-EGF7-10: 1D NMR spectra of EGF7-10 in different buffers with different pH levels, each indicated by their colour coding.

However, a fresh ^{15}N-EGF7-10 sample was prepared in a MES buffer for later use in calcium cation binding assays. This sample yielded a well-dispersed set of proton peaks in the 1D NMR (figure 5.26 A), and well-resolved amide signals on the HSQC, which were likewise wee-dispersed (figure 5.26 B). However, the NOESY shown in figure 4.35 C was still relatively weak compared to the one obtained for EGF5-7 in figure 4.31 B; although the presence of through-space cross peaks is still able to be visualized. Nonetheless, through-space cross peaks between side-chain methyl groups and backbone amides, consistent with folded protein, are visible in this spectrum. This sample was judged to be of sufficient quality for use in the calcium-binding studies described in section. 5.7.

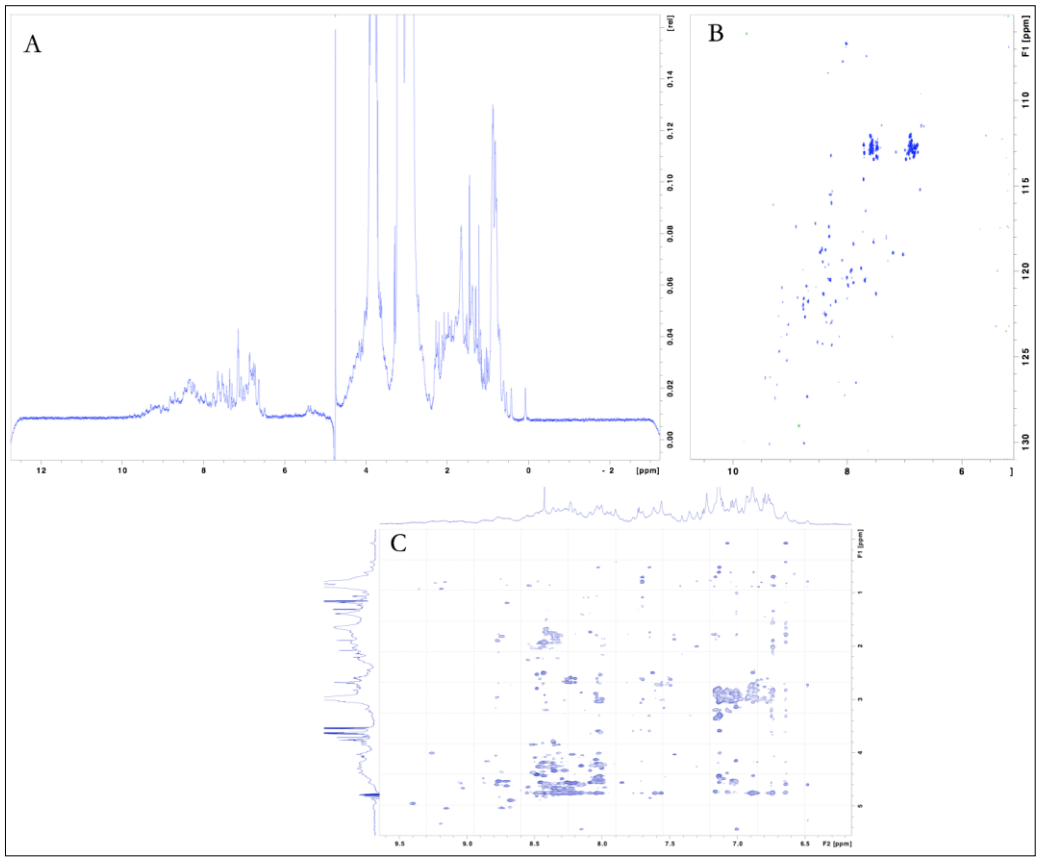


Figure 5.26. ^{1}H, HSQC and NOESY NMR of ^{15}N-EGF7-10: A. 1D NMR spectra of 32 μM EGF7-10 (20 mM MES, 75 mM NaCl, pH 6.55). B. HSQC spectra obtained over 60-minutes. C. 2D NOESY on the ^{1}H - ^{1}H plane of EGF7-10. All NMR experiments were carried out at 25 ºC.

5.4.2.5 Crystallography

No potential crystals were observed for the EGF7-10 sample, indicating it would require further optimization for crystallization. After initial crystallization trials with a 12 mg/ml sample of EGF5-7 failed to yield any protein crystals (as determined by synchrotron analysis, data not shown), more concentrated EGF5-7 samples (25 mg/ml) were screened, with three wells yielding potential crystals (figure 5.27). When cryoprotectant was added to the well shown in figure 5.27 A, salt crystals rapidly formed indicating the cryoprotectant was unsuitable for mixing with the buffer in the well, and rendering the sample useless (unfortunately, due to

delays from the EPPF facility over several months, the well partially dried out, making it difficult to remove a sample of the contents that could have been used to first check this procedure). The hexagonal crystal in the well shown in figure 5.27 B was successfully loaded into a loop and sent to the synchrotron for analysis, revealing the presence of what appeared to be small molecules, not protein (data not shown). In attempting to load into a loop the potential crystal shown in figure 5.27 C, it crumbled into powdery pieces in a way that indicated it might have been a protein crystal (personal communication, Atlanta Cook, June 2017). Based on the encouraging results for two of the three potential crystals, a set of screening conditions was set up around the contents of the wells from figure 5.27 A and C. However, after three months, whilst precipitation occurred in over 50% of the wells, no crystals were formed, indicating that further optimization would be required.

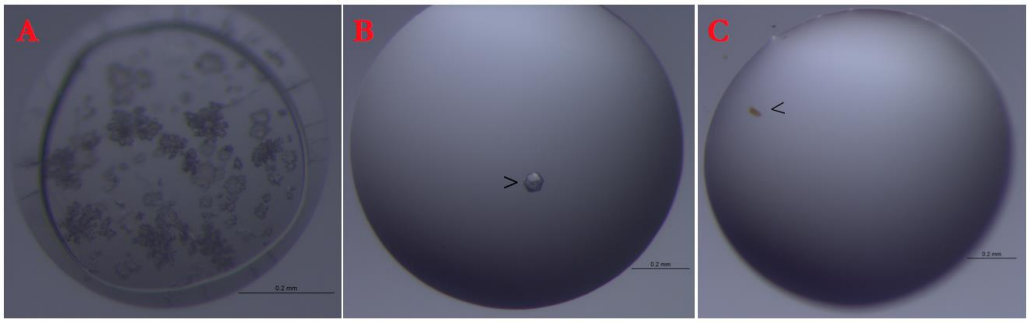


Figure 5.27. Second Crystal Screen of EGF5-7: A. A cluster of fractal-like crystals. Conditions- 40% v/v ethanol/5% w/v PEG 1000 (precipitant), no salt, 0.1 M phosphate/citrate pH 4.2. B. A single hexagonal crystal. Conditions- 1.8 M ammonium phosphate monobasic, 0.1 M Sodium acetate trihydrate pH 4.6. C. A single brownish crystal shaped like a triangular log. Conditions- 4.3 M NaCl (precipitant), no salt, 0.1 M Na HEPES pH 7.5.

5.4.2.6 Homology-Based Modelling and SAXS Analysis

Using SWISS-Model (and confirming overall structural similarity with another modelling program, I-Tasser), homology-based models for EGFs 5-7 and 7-10 were generated with the final selection based on the best Global Model Quality Estimation (GMQE) and QMEAN scores (figure 5.28 A and B). For both proteins, the GMQE scores (a measure of quality estimation combining qualities of target template alignment, in which higher numbers indicating higher quality; numbers between 0 and 1) reflected a mediocre level of model quality when compared to the library of PDB structures used to assess the model against. Likewise, the QMEAN scores (a type of composite scoring based on various geometrical properties to give an absolute quality estimate on the basis of a single molecule model) of both proteins reflect a model of potentially low quality; any QMEAN is said to be of unreliable quality if the value is below -4.0. Due to the various primary sequences that can give rise to EGF-like domains, the alignments and comparisons to libraries of proteins used by SWISS-Model may have had a significant impact on the scores retrieved. Due to this, a basic comparison between the CD data from figure 5.18 (for EGF5-7) and the modelling data was carried out to assess the amount of secondary structure each gave (table 4). This revealed that the CD data overall best matched the SWISS-Model results rather than I-Tasser, showing each to have approximately the same level of α-helix and β-strand in terms of percentage of residues analysed, with the CD data estimating slightly higher levels of secondary structure than the modelling predicted (however this may be due to the model only being able to cover 84% of the protein sequence). The I-Tasser model data gave a more accurate match of α-helix, but was significantly lower for the estimated amount of β-strand, with almost half as much predicted.

Both protein models of EGF5-7 and 7-10 contained β-strand in each linker region between the EGF-like domains. Characteristic anti-parallel β-strand were predicted to lie within these domains. Both proteins were predicted to be elongated rather

than globular, as also observed in many other proteins comprised of in-tandem EGF-like domains. However, it should be mentioned that the program assumes that the disulfide bonds are in the regular order as found in most EGF-like domains; that is Cys1-Cys3, Cys2-Cys4 and Cys5-Cys6. If this assumption turns out to be false, it degrades the usefulness of the model, particularly since it impacts on the predictions of domain boundaries. Nonetheless, the model generated here could be useful for estimating the surface positions of any molecular partner interactions mapped by chemical shift perturbations (in lieu of a high-resolution crystal or NMR structure).

Next, SAXS-based analysis was carried out on a sample of EGF5-7 (0.5 mM protein in 20 mM sodium acetate pH 4 and 75 mM NaCl). After the collection of data, the homology based model for EGF5-7 from figure 4.38 A was used to fit the SAXS data using the program CRYSOL (Appendix figure A1) to yield a χ^2 value of 1.0, which lies within the acceptable range of 0.9 and 1.2 (Franke, Jefferies, & Svergun, 2015). Due to the goodness of fit between experimental and collected data, an *ab initio* approach was taken using the homology-based model generated in figure 4.38 A using the modelling program DAMMIF, and the resultant models superimposed over each other (figure 5.28 C and D). The overall SAXS model shows an elongated, tick-shaped molecule and an overall good fit when superimposed with the space-filling model of EGF5-7 (figure 5.28 C). Interestingly, the original structure derived from homology-based modelling best fits into the *ab initio* SAXS model when linker flexibility allows for increased elongation between EGF's 5 and 6 as demonstrated in figure 5.28 D.

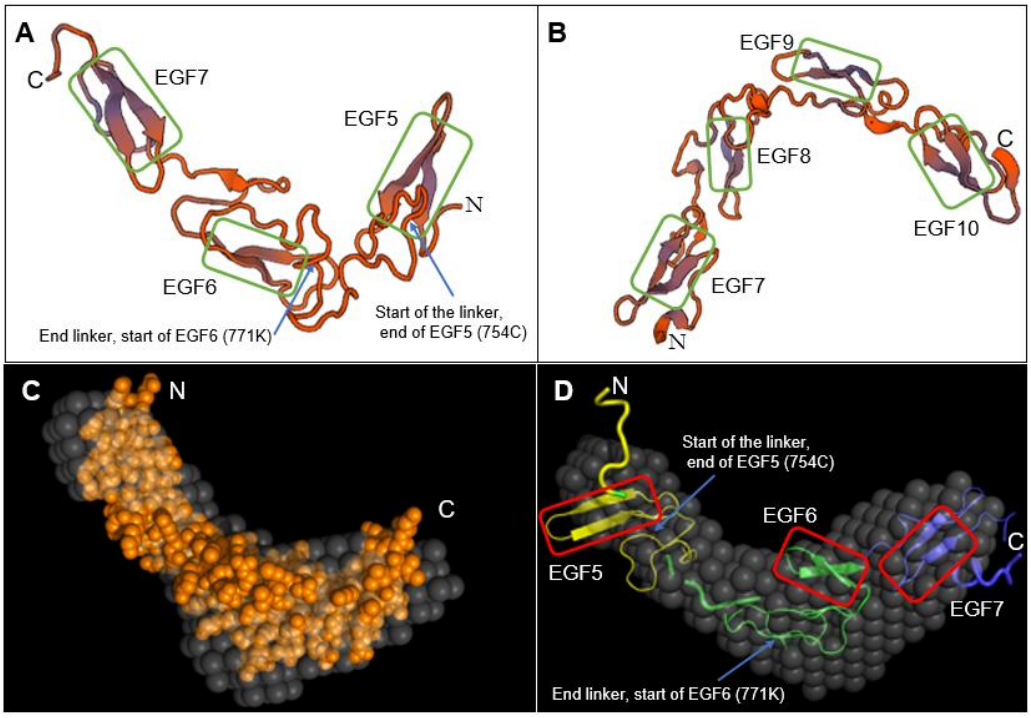


Figure 5.28. Homology Modelling and SAXS Models: A. The model generated for EGF5-7. The N-terminal (N) domain starts with Ser 723 of EGF5, ending at Ile 753, before the start of the linker commences at Cys 754. Each EGF-like domain's anti-parallel β-strand is highlighted by the green boxes. The C-terminus (C) ends at Lys 846. 69% of the sequence was given computational coverage. The GMQE is 0.40 and the QMEAN score is -5.57. The model generated is based on homology with Neurogenic Notch Homologue Protein1 (5FMA in the PDB). B. The model generated for EGF7-10. The N-terminus starts with Tyr 818 of EGF7, and ends with Asn 1005 at the C-terminus. 88% of the sequence was given computational coverage. The GMQE is 0.53 and the QMEAN score is -4.78. The model generated is based on homology with Neurogenic Notch Homologue Protein1 (5FMA in the PDB). C. The average fitted *ab initio* model of EGF5-7 produced via SAXS modelling. Grey spheres are from the analysed SAXS data whilst the orange spheres are the fitted-in data from the homology-based model derived from (A), with adjustments made for the potential flexibility of the long linker region. D. Rigid body modelling of the homology-model structure of EGF5-7 from (A), with adjustments made for the potential flexibility of the long linker region with fitting into the SAXS-derived data via special alignment.

	Swiss model data	I-Tasser data	CD data
α-helix	3.03%	5%	4.8%
β-strand	31.06%	17.85%	35.3%

Table 4. Computational Modelling and CD Data Comparison. Data given are for EGF5-7.

5.4.2.7 Assessment of Free Cysteines and Disulfide Mapping

Despite the PfRipr proteins having been produced in *P. pastoris*, a eukaryotic expression system (preferably, in numerous ways, to the use of a bacterial host that would necessitate re-folding in many cases), it was important to assess the potential that the cysteines found in the linker were oxidized as a disulfide bridge as opposed to being reduced and free cysteine residues. Further, a eukaryotic-expression system is not always guaranteed to produce correctly-folded recombinant protein; therefore, it becomes important to still assess these proteins for evidence of correct folding (covered in more detail in (Gasser, *et al.*, 2007)). A sample of EGF5-7 was used to assess the presence of free cysteine residues by exposing a third of the sample to NEM (which would associate with free thiols), a third to a reducing agent, then NEM, and a third as a control. The results in figure 5.29 show that all cysteines form disulfide bonds based on the fact that no size change is detected from the control sample. However, due to the paucity of molecular weight increase when EGF5-7 was mixed with NEM and TCEP (figure 5.29 C), this suggests a failure to reduce the protein with the TCEP used.

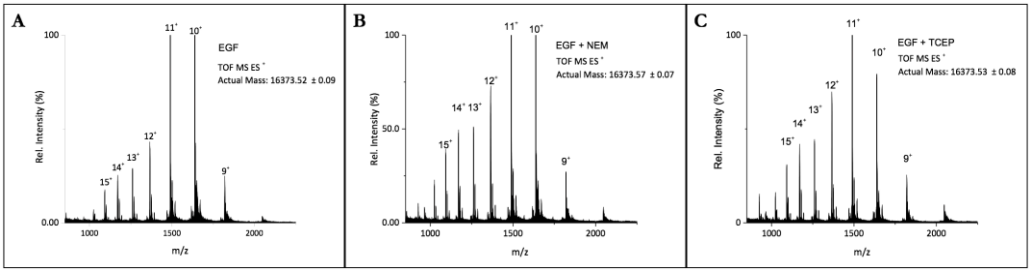


Figure 5.29. MS-ESI Assessment of Free Cysteines in EGF5-7: A. Native EGF5-7. B. EGF5-7 + NEM. C. EGF5-7 reduced with TCEP + NEM.

Pepsin was used, initially, to enzymatically process EGF5-7 prior to disulfide mapping using MS, however the results obtained suggested incomplete digestion (data not shown). Trypsin, likewise, produced an incomplete enzymatic digest. Trypsin preferentially works at a pH of 6.8, potentially explaining its poor activity when EGF5-7 was presented as substrate in an acetate buffer at pH 4. A buffer-

exchange was attempted on a small size-exclusion column, but this did not appear to go to completion based on the final pH of the sample. A mini dialyzer was chosen for efficient buffer exchange so that the tryptic digest could be carried out again; however, at the time of writing this experiment is on-going.

5.5 1D, 2D and Triple-Resonance NMR Analysis of $^{15}N/^{13}C$-Labeled EGF5-7

5.5.1 Methods: 1D, 2D and Triple-Resonance NMR Analysis of $^{15}N/^{13}C$-Labeled EGF5-7

NMR was carried out as per section 4.4.1.5 at the University of Edinburgh using the 600 MHz Bruker AVANCE 14.1 – Tesla NMR spectrometer. The 800 MHz spectrometer was used to collect data for T_1/T_2 relaxation and heteronuclear ^{1}H-^{15}N NOE experiments (all relaxation experiments carried out at 25 ºC). CCPNMR Analysis software version 2.4.2 (http://www.ccpn.ac.uk/v2-software/software/analysis) was used. Dr. Brian Smith at the University of Glasgow provided training and help with the data analysis. The suite of experiments carried out for 3D structure determination included- ^{15}N HSQC, ^{13}C HSQC, CBCA(CO)NH, CBCANH, HBHANH, HBHA(CO)NH, HN(CA)CO, HNCO, NOESY HSQC NH, NOESY HSQC ^{13}C, HCC(CO)NH, HCCH, HBCBGCDHD, HBCBCGCDCEHE, HSQC CTET, and ^{15}N-edited TOCSY HSQC. Experiments were first carried out at 25 ºC then repeated at 40 ºC. An overview of resonance assignment and assignment of backbone resonances can be found in sections 2.3.4.3 and 2.3.4.4, respectively, (Makou, 2013).

Relaxation data plotting using Prism was carried out by William Godfrey at the Institute for Molecular Bioscience, University of Queensland.

5.5.2 Results: 1D, 2D and Triple-Resonance NMR Analysis of $^{15}N/^{13}C$-Labeled EGF5-7

For 3D structure determination by NMR of a sizeable protein, a plethora of 2D and 3D experiments must be carried out on a $^{15}N/^{13}C$-labeled sample. The $^{1}H,^{15}N$-HSQC serves as the reference spectrum. Each cross peak in the HSQC that corresponds to a backbone amide group, must be carefully picked by the software with manual intervention, and then catalogued using the software. The produces a table which contains chemical shifts in both the nitrogen dimension and in the ^{1}H dimension for each backbone amide and the main initial task is to assign each cross-peak to a sequence-specific amino acid residue. The results of the initial ^{15}N-HSQC of $^{15}N/^{13}C$-EGF5-7 are shown in figure 5.30. Approximately 120 peaks were visible, excluding significantly overlapped peaks and the peaks corresponding to side chain amide groups of asparagine and glutamine residues. No significant differences were noted between the $^{15}N/^{13}C$-labeled sample and the dll sample that was previously made.

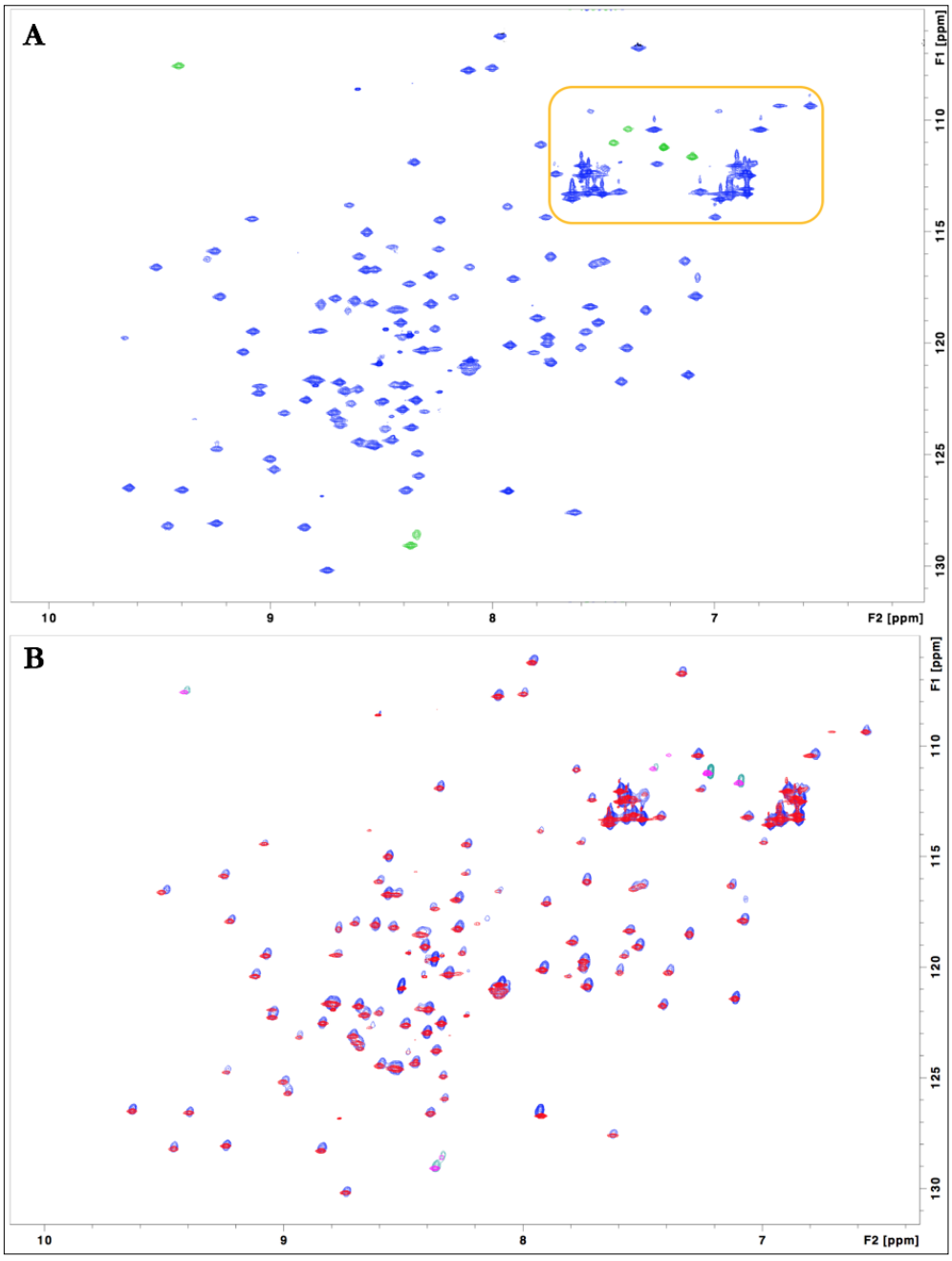


Figure 5.30. ^{15}N-HSQC of ^{15}N/^{13}C-EGF5-7 at 25 °C: A. The protein was at a concentration of 0.5 mM and in a buffer comprised of 20 mM sodium acetate pH 4 + 45 mM NaCl + 10% D_2O. Folded peaks are shown in green and the orange box highlights the location of the Asn/Gln NH_2 peaks. B. An overlay with the HSQC spectra obtained from the dll-EGF5-7 sample (red spectra is ^{15}N/^{13}C-EGF5-7, blue is dll-EGF5-7).

After using the automated peak picking tool, manually removing the erroneously picked peaks, and finding hidden peaks (validated by confirming signals present in each of the CBCA(CO)NH, CBCANH, HNCO and HN(CA)CO spectra), the total peak count was at 138. The $^{15}N/^{13}C$-EGF5-7 protein is comprised of 144 amino acids including the Ala-Gly-Ile-Pro cloning artefact at the N-terminus (a proportion of the protein molecules will have an additional Glu-Ala in front of this), with four proline residues that will not produce an HSQC signal due to them not possessing an amide group. Therefore, one would expect to be able to count 139 peaks if the N-terminal Ala residue is undergoing intermediate exchange; otherwise 140 peaks should be seen. Given that 138 peaks were located, this indicates another residue may be in intermediate exchange, or the C-terminal Glu residue is missing from the protein. Further, to complete the peak identification, data from repeating the aforementioned NMR experiments at 40 °C were used, as this resulted in better peak separation. For this a temperature scan was taken from 10 – 40 °C to track the peaks and transfer the picked peaks accordingly (figure 5.31). At the higher temperature there was a significant improvement in the quality of the spectra, with improved resolution of peaks that were clustered together or overlapping. This led to a significant improvement in the level of accurate peak picking that would facilitate the in-depth analysis of the type of amino acid most of these peaks represented. Once a peak was picked, the atoms of the amide group of the residue were assigned.

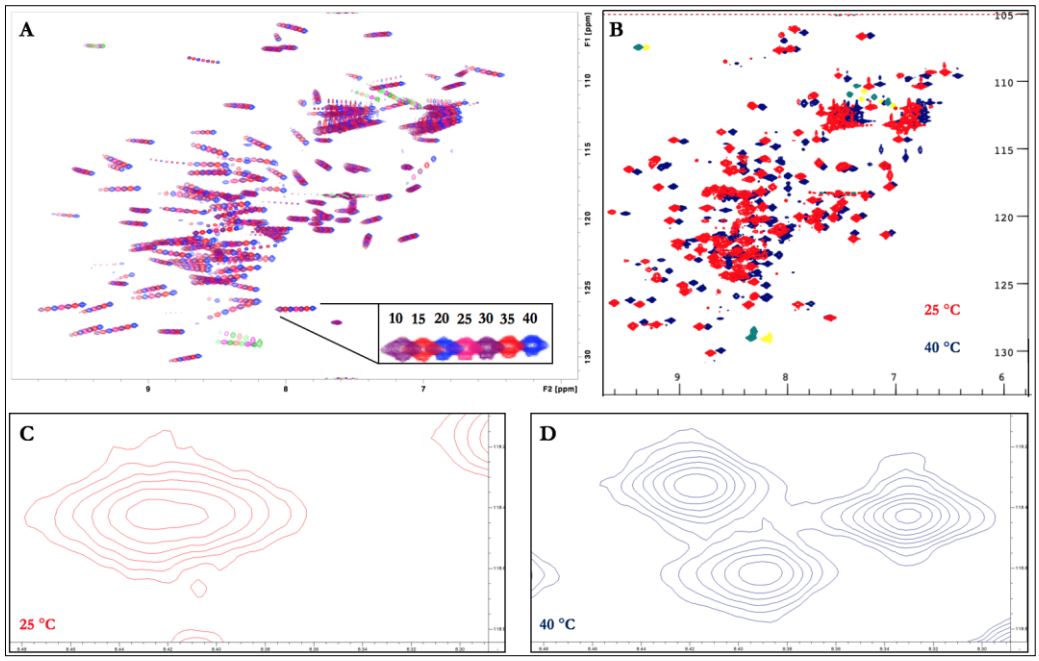


Figure 5.31. ^{15}N-HSQC Temperature Scan for ^{15}N/^{13}C-EGF5-7: A. An overlay of HSQC spectra from 10 °C to 40 °C, going up in 5 °C increments (the inlaid box show which peaks were acquired at which temperature in °C). B. The overlay of the HSQC spectra for samples tested at 25 °C and 40 °C. C. A zoomed-in example of what appeared to be a single peak from the HSQC acquired at 25 °C. D. The peak in (C) contained a group of three overlapping peaks that was revealed in the HSQC run at 40 °C. Spectra were acquired over the course of one-hour for each temperature point.

In order to garner further information about each peak, two complementary pairs of experiments were used (CBCANH + CBCA(CO)NH, and HNCO + HN(CA)CO), providing signals for the Cα, Cβ, and carbonyl atoms (the carbon of the CO group). An example of this is shown in figure 5.32. The base amide signal serves as the "i" amino acid residue (that is, the residue of interest), with the CBCANH experiment revealing its Cβ and Cα atoms. Where a peak appears in both the CBCANH and CBCA(CO)NH spectra, it will overlap, indicating this is the Cβ and Cα atoms of the previous amino acid (i-1). The HN(CA)CO and HNCO spectra combined gave rise to three peaks; a larger single peak in the HNCO spectra representing the i-1 ^{13}C(O), and two peaks in the HN(CA)CO,

with the smaller overlapping with the HNCO peak and the other belonging to the $^{13}C(O)$ of the i-amino acid.

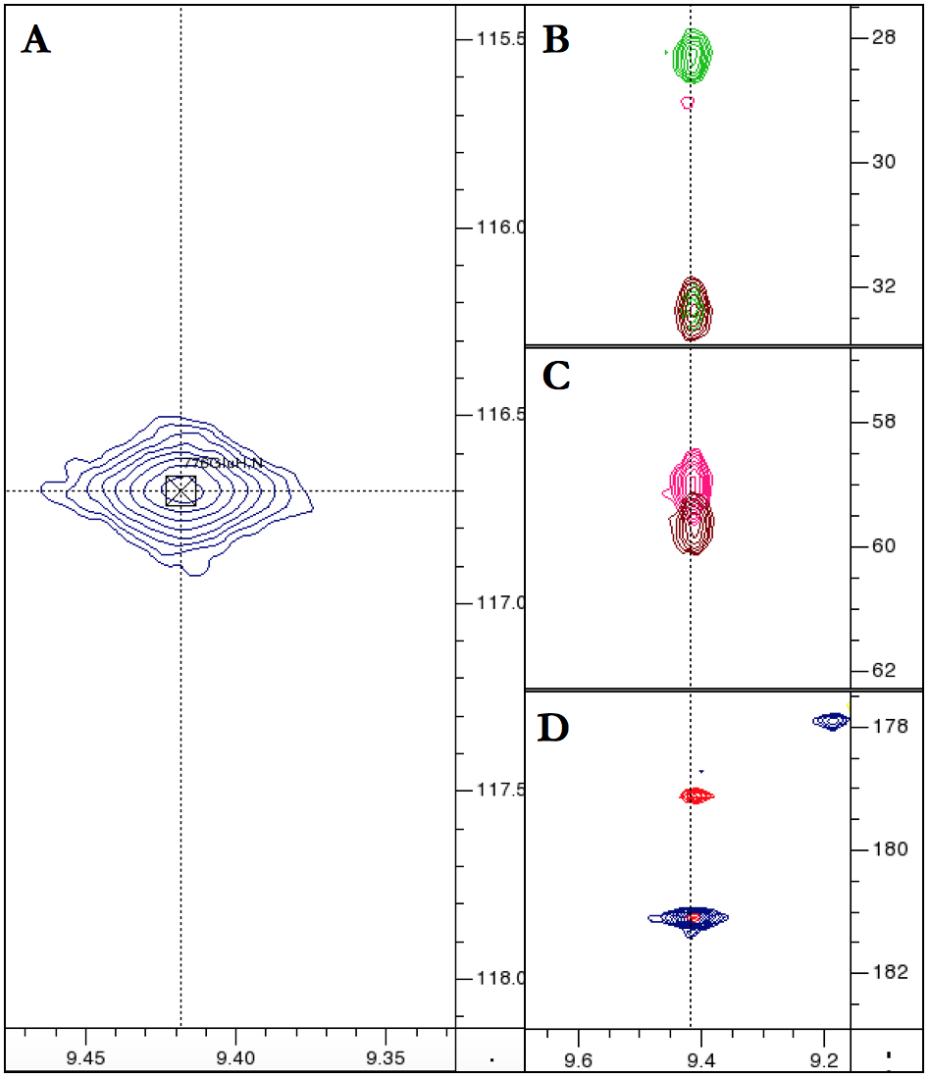


Figure 5.32. Cα, Cβ and C(O) Assignment: A. The base amide peak from the HSQC experiment shown in figure 4.41 B (40 °C data set). Navigation from this peak to the carbon dimension reveals the peaks seen in B – D. This base amide signal is amino acid "i". B. A pair of Cβ carbons with the brown peak belonging to the results of the CBCA(CO)NH data (therefore representing the i-1 amino acids' Cβ) and the green peak belonging to the CBCANH data (therefore, this Cβ belonging to the i-amino acid). C. A pair of Cα carbons with the i-1 Cα (brown) and the i Cα (pink). D. A pair of C(O) peaks; i-1 in blue and i in red.

After peaks in the triple resonance NMR spectra described above were picked and the atom resonance and type assigned, the next step was to commence the backbone assignment by either going in the i-1 or i+1 direction. This involved setting up "strips" corresponding to amide route resonances, for the linking of sequential residues using the ^{13}C experiments. This "walk" between residues allows for increased confidence with assignments using the CCPNMR Analysis tool. Shown in figure 5.33 is an example of one such walk, with assigned residues K775 – N779 shown.

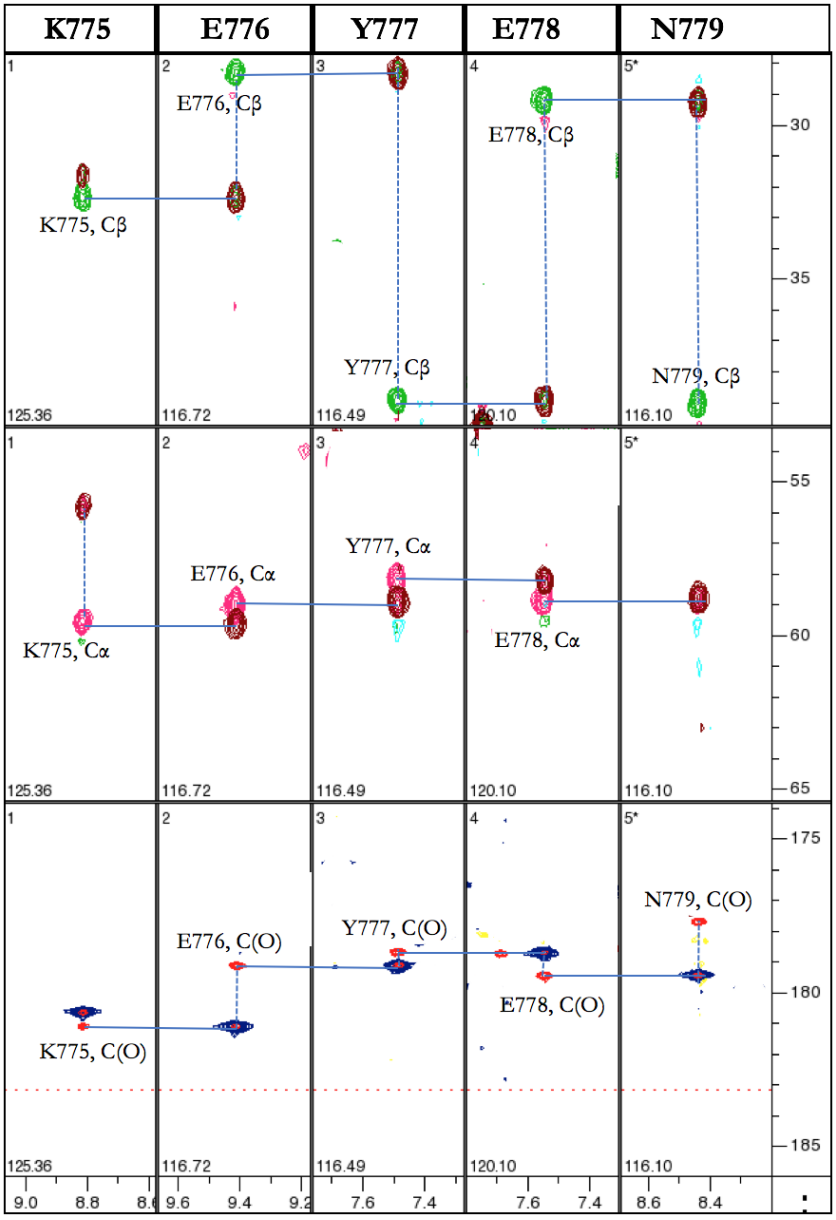
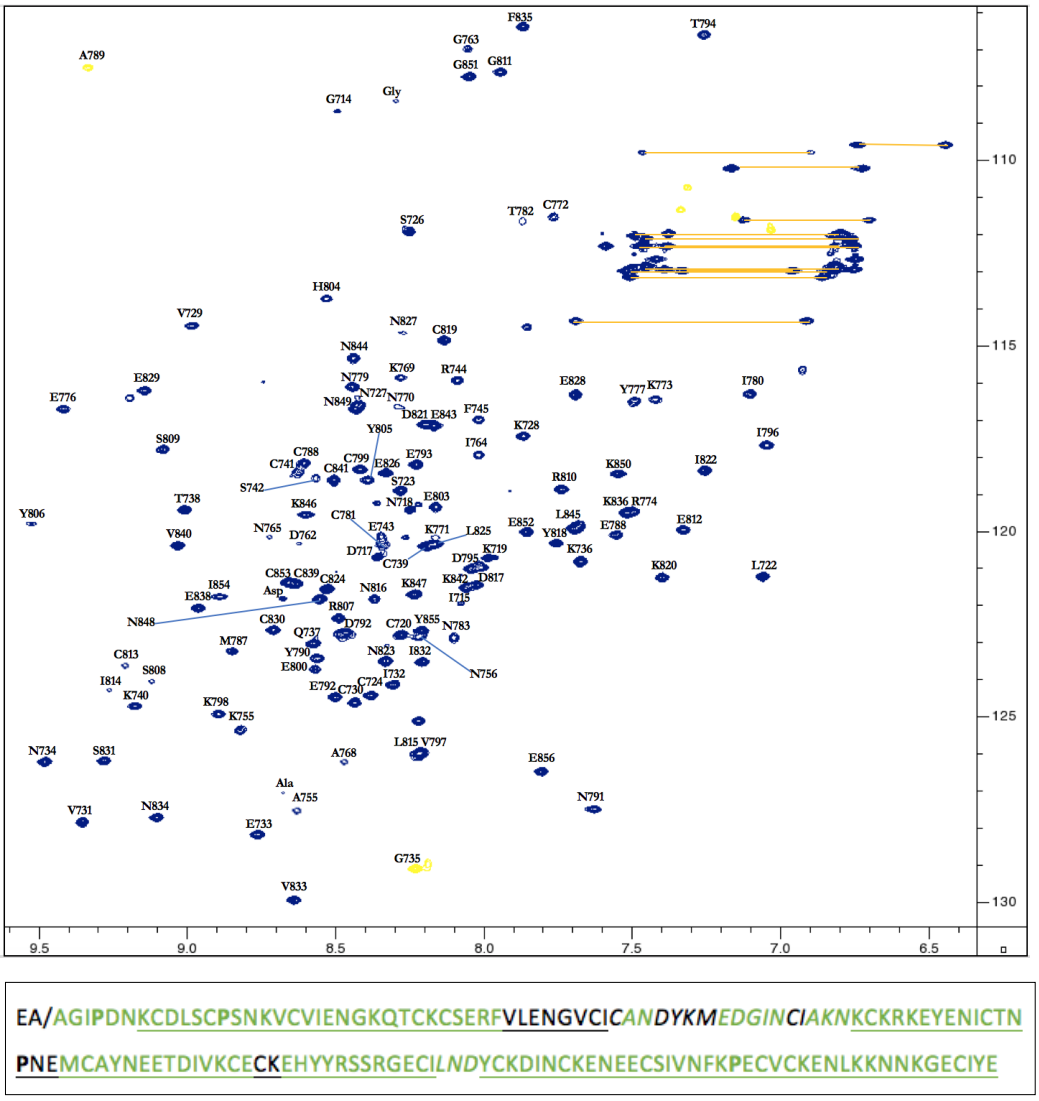


Figure 5.33. Sequential Strips Walking Along the Backbone of K775 – N779: The strips have been placed in sequential order and correspond to residues K775 – N799 of PfRipr. Each Cβ, Cα and C(O) atom has been labelled for their respective residue. Solid lines join the i-amino acid with the i+1 residues, whilst the dotted lines connect the i-1 residue to the i-amino acid within the same strip.

Regarding the data acquired at 40 °C, this method was utilized to attain 86% of the backbone assignment (up from 70% with the original 25 °C data), with

confirmation of various difficult to assign residues made via the additional incorporation of HBHA(CO)NH, HBHANH, HC(CCO)NH and NH-NOESY experiment spectra. These experiments allowed for an investigation of the side-chain protons, increasing confidence as to the type of amino acid residue assigned to a given HSQC peak. The chart shown in figure 5.34 shows the primary sequence of EGF5-7 indicating the extent of NH and NH Cα and Cβ assignments along with HSQC showing assignments for 124 out of 144 amino acids residues. Note that full coverage of EGF7 was achieved.



EA/AGIPDNKCDLSCPSNKVCVIENGKQTCKCSERFVLENGVCIC*ANDYKM**EDGIN*CI*AKN*KCKRKEYENICTN

PN**E**MCAYNEETDIVKCE**CK**EHYYRSSRGECI*LND*YCKDINCKENEECSIVNFK**P**ECVCKENLKKNNKGECIYE

Figure 5.34. Assigned ^{1}H-^{15}N-HSQC Spectrum of EGF5-7: Sidechains of NH resonances for Gln and Asn residues are connected by orange lines. Folded peaks are shown in yellow. Where the amino acid type is assigned, but not the location in the protein, the 3-letter code is used. Below the HSQC is the assignment log; green letters are assigned residues, black are unassigned, prolines are in bold, linker regions are italicised, and each EGF-like domain is underlined.

All of the unaccounted for residues are predicted to be in flexible loop regions of the protein (based on the SWISS-Model generated structure shown in figure 4.38 A), with the exception of C766 and I767 which are predicted to be at the end of a putative β-strand in the long linker region between EGFs 5 and 6. Due to numerous delays out of the control of the author, completion of the 3D structure

of EGF5-7 was unable to be carried out in the time allowed. However, once the side-chains become assigned, the missing residues may be teased out. If this does not occur, then analysis of the data from the experiments carried out at 10 °C may help reveal additional connecting peaks in the carbon dimension experiments.

However, given the level of backbone assignment a simple histogram could be generated for the mobility of each known amino acid (specifically, the proton delta chemical shift) during the temperature shift between 25 and 40 °C, based on the data in figure 5.31 B (see figure 5.35 for the results of the plot). The largest peak shift was seen with V797 in EGF6 at just under 0.2 ppm, predicted to be in a flexible loop connecting the anti-parrallel β-strands. No discernable shift was seen in N791, which is predicted to be at the tip of the first anti-parrallel β-sheet. Using a cut-off of 0.1 ppm, the EGF-like domain that had the lowest percentage of significant shift during the temperature increase (significant here defined as a shift over 0.1 ppm) was EGF6, with 60% of the assigned residues not undergoing a signficant shift. EGF5 had 60% of the assigned residues undergoing significant chemical shifts, and EGF7 had slightly over 55% undergoing a significant chemical shift. This may suggest that EGF6 is slightly more rigid than the flanking domains.

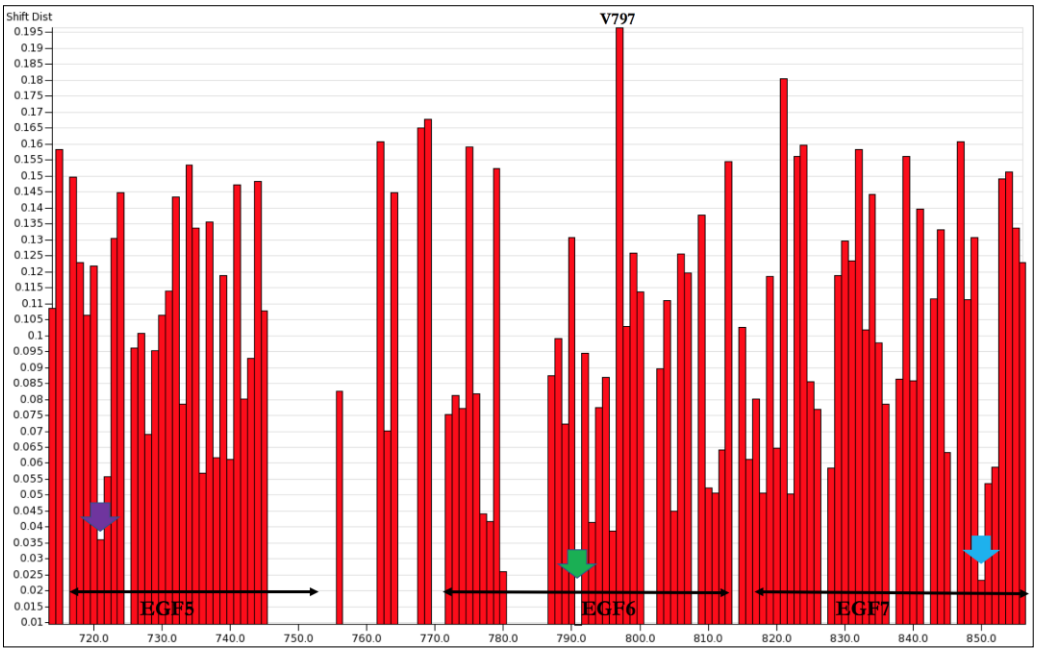


Figure 5.35. Residue Shifts From 25 – 40 °C: The demarcation of each EGF-like domain is given at the bottom of the histogram. Each arrow highlights the residue that was least affected by the increase in temperature; purple is D721, green is N791, and blue is K850. The largest shift is highlighted was V797. Shift distance is in ppm. Histogram made with CCPNMR Analysis.

To further understand the dynamics of EGF5-7, T_1 (spin-lattice), T_2 (spin-spin), and heteronuclear (^{1}H-^{15}N) NOE relaxation experiments were carried out at two magnetic field strengths; 600 MHz and 800 MHz (so as to create a comparison of the data sets). An analysis of these parameters should help identify and differentiate between mobile and rigid amino acids residues, which can be discussed in the context of the homology-based structual model of the protein. The results are summarised in figures 5.36 – 5.38.

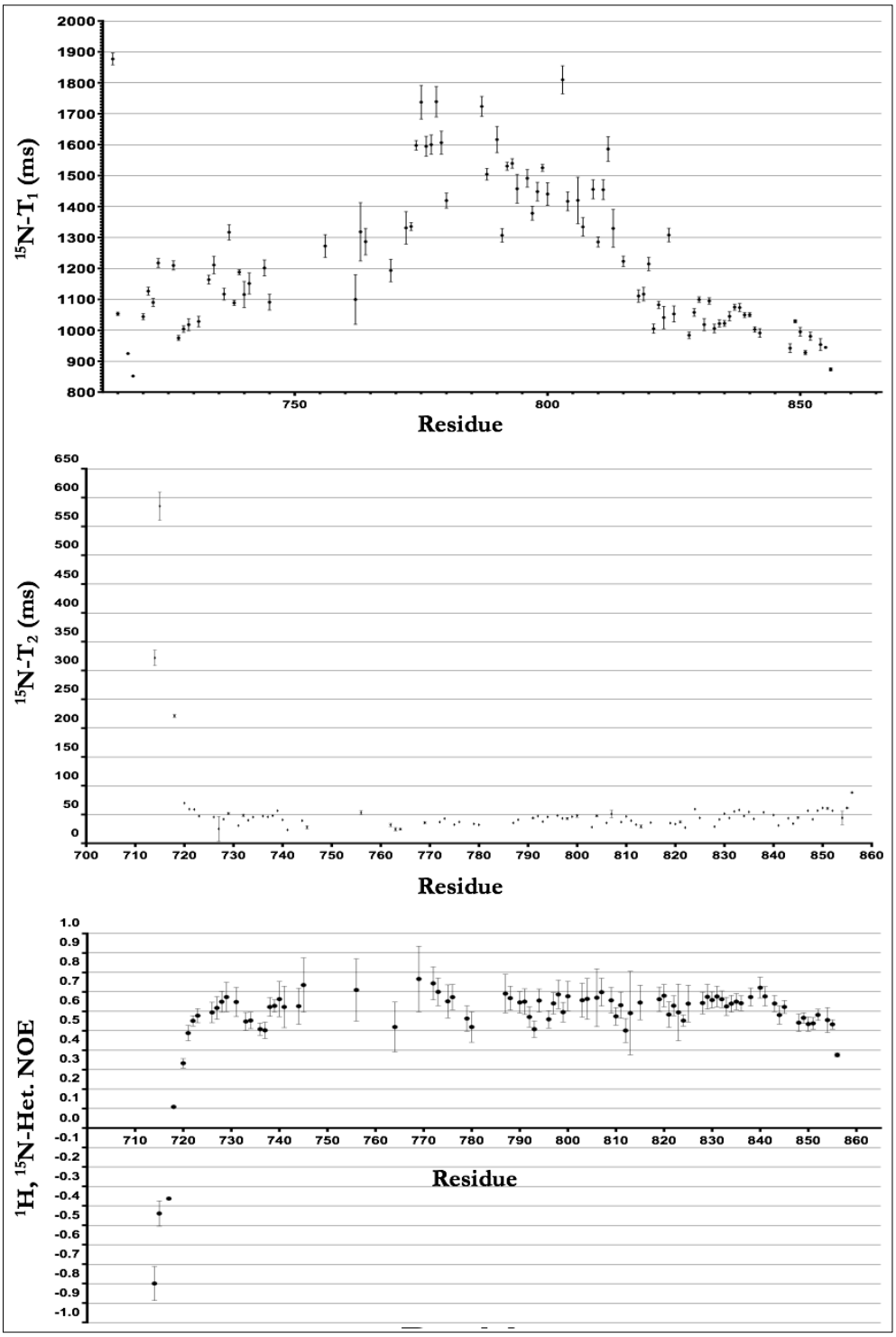


Figure 5.36. ^{15}N Relaxation Data for ^{15}N/^{13}C-EGF5-7 (800 MHz): Plots are shown for the relaxation times measured in milliseconds (ms) for the ^{15}N spin-lattice (T_1) and ^{15}N spin-spin (T_2) experiments, in addition to the heteronuclear (^{1}H, ^{15}N) NOE values, plotted against residue number. Error bars are given in standard deviation.

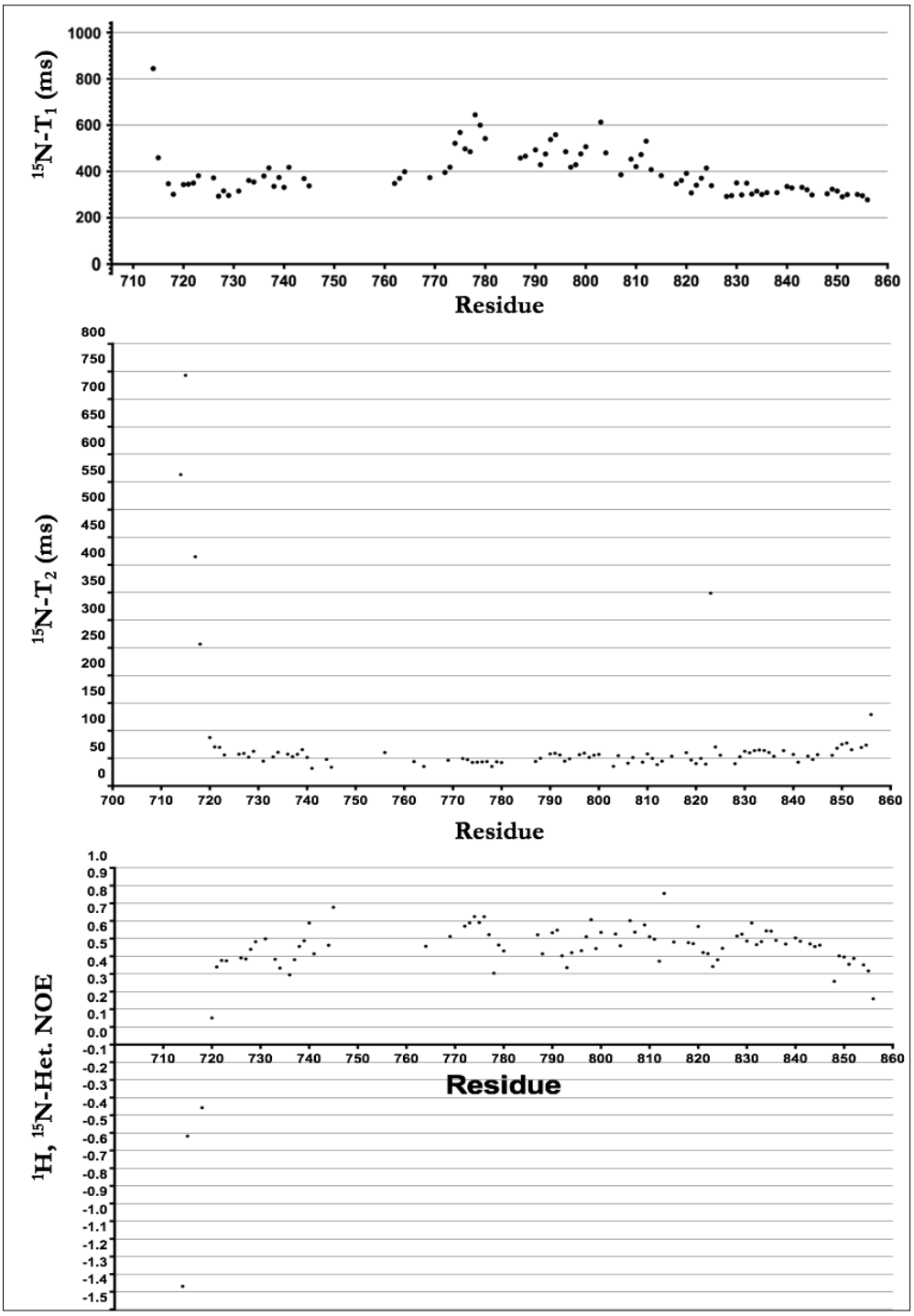


Figure 5.37. ^{15}N Relaxation Data for ^{15}N/^{13}C-EGF5-7 (600 MHz): Plots are shown for the relaxation times measured in ms for the ^{15}N spin-lattice (T_1) and ^{15}N spin-spin (T_2) experiments, in addition to the heteronuclear (^{1}H, ^{15}N) NOE values, plotted against residue number. No error bars given due to an error with CCPNMR Analysis.

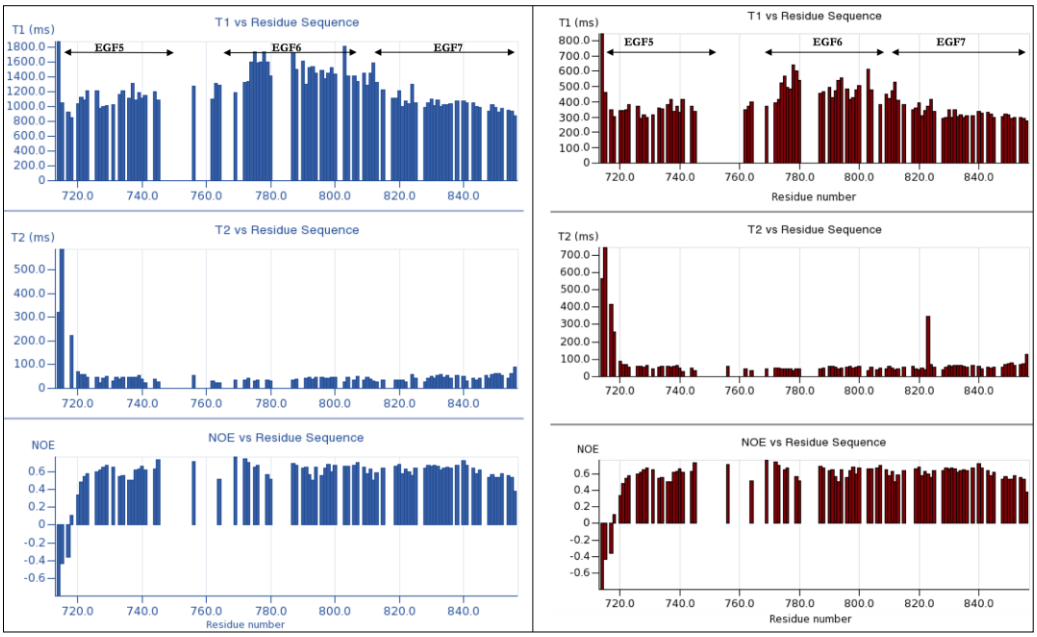


Figure 5.38. Histogram of ^{15}N Relaxation Data for ^{15}N/^{13}C-EGF5-7 (800 MHz V 600 MHz): Plots are shown for the relaxation times measured in ms for the ^{15}N spin-lattice (T_1) and ^{15}N spin-spin (T_2) experiments, in addition to the heteronuclear (^{1}H, ^{15}N) NOE values, plotted against residue number. Data in blue is that from the 800 MHz spectrometer whilst the maroon data is from the 600 MHz spectrometer. Each EGF-like domain boundaries are shown atop the figure.

Data collected from both fields shared the overall same pattern; more disorder in the first four residues that comprise the floppy N-terminal region (N714, K715, C716, and D717) as reflected by the long T_1 and T_2 values combined with the negative heteronuclear NOE values, relatively uniform NOE levels with a drop in the C-terminus (indicative of some flexibility), and an overall pattern of a high T_1 to T_2 ratio indicative of structural rigidity, with the T_1 values highest in the second EGF-like domain of EGF5-7. The one anomaly is the increased T_2 signal in the 600 MHz data for residue N823. The anomalous T_2 value (in the 600 MHz data) for residue N823 is likely to be an artefact of the fitting process.

More specifically, long T_1 values that are correlated with short T_2 measurements indicate molecular rigidity, however, these generally are scaled together; the residues from 770 to 810 (of EGF6) have longer T_1 values than the residues in EGF5 or 7, yet the T_2 values remain the same overall. Given the protein size of

16.52 kDa, one might expect the T_2 values to be around 100 ms for these residues if there were scaled proportionally to the T_1 increases. This indicates that not only is the region inflexible but that it is likely self-associating with other molecules of EGF5-7 in solution. Given that the T_1 ratio increase is not present in the residues of the other EGF domains, EGF5-7 is probably undergoing intermolecular association via the central EGF domain whilst the protein as a whole does not behave as one entity. Given that this dimerization was not seen on gel filtration analysis, it is likely this association (if truly occurring) is due to the relatively high protein concentration of 0.5 mM. The combination of flexible regions of the protein and intermolecular association may also explain why crystallography continued to fail despite testing a variety of protein concentrations and crystal buffer conditions.

With respect to the overall uniformity of the NOE, the mean measurement of non-terminal residues analysed was 0.56 (0.51 for the 600 MHz data), indicative of a relatively rigid protein (the closer the value to 1, the more rigid, whilst the inverse is true). The residues that significantly dropped below this value in the NOE (aside from the aforementioned N- and C-terminal residues) include: K736, Q737 and C739 in EGF5, G763 in the long linker between EGFs 5 and 6, and E793 in EGF6, K842 and G843 in the C-terminus of EGF7. By looking at the zoom-in of the previously established SWISS-Model predicted structure of EGF5-7, all of these residues are found in loop regions of the protein, which would make sense if they are more flexible relative to the rest of the protein (figure 5.39).

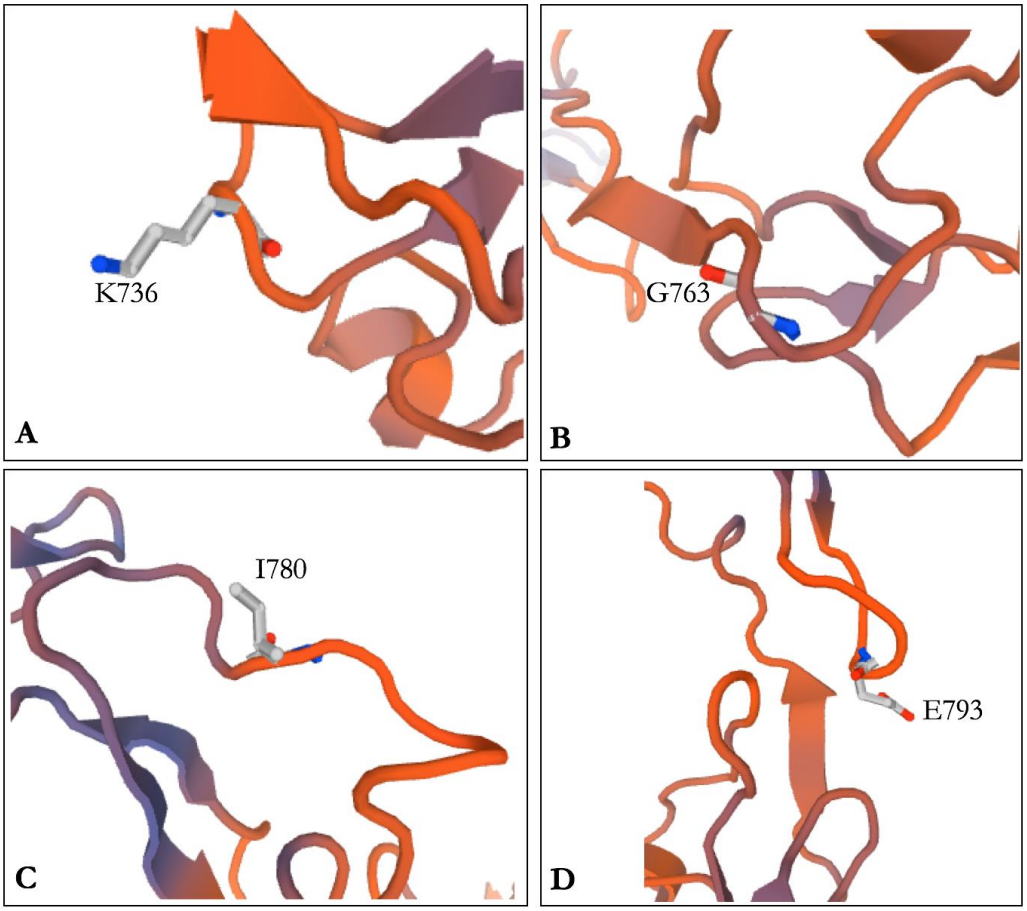


Figure 5.39. Location of Residues with Low Heteronuclear NOE Values: The above model shows the predicted structural regions of some residues identified as having a below average heteronuclear NOE score relative to the rest of the residues comprising EGF5-7 (excluding N- and C-terminal flexible regions. All are predicted to be found in loop regions. Model made with SWISS-Model.

Unfortunately, lack of time prevented follow up studies including the total correlation calculations or measurement of residual dipolar coupling (RDC). However, an NMR-based assessment of the stability of $^{15}N/^{13}C$-EGF5-7 over time was carried out using 1D and 2D NMR techniques. Figure 5.40 display the 1D spectra obtained from three points in time over a ~9-month period, where the sample had spent ~40 days at 25 °C, 10 days at 40 °C and the rest of the time at 4 °C. No anomalies are found in the amide region, but there are two minor additional peaks that appear in the methyl region at around 3.8 and 1.8 ppm after

the 4.5 months. However, these peaks were very slightly present on the shoulders of the main peaks they are found on even at the beginning of the experiment and do not overtly look like the products of any protein degradation. Additional evidence for a dearth of sample degradation was a lack of relative loss of intensity of signals given.

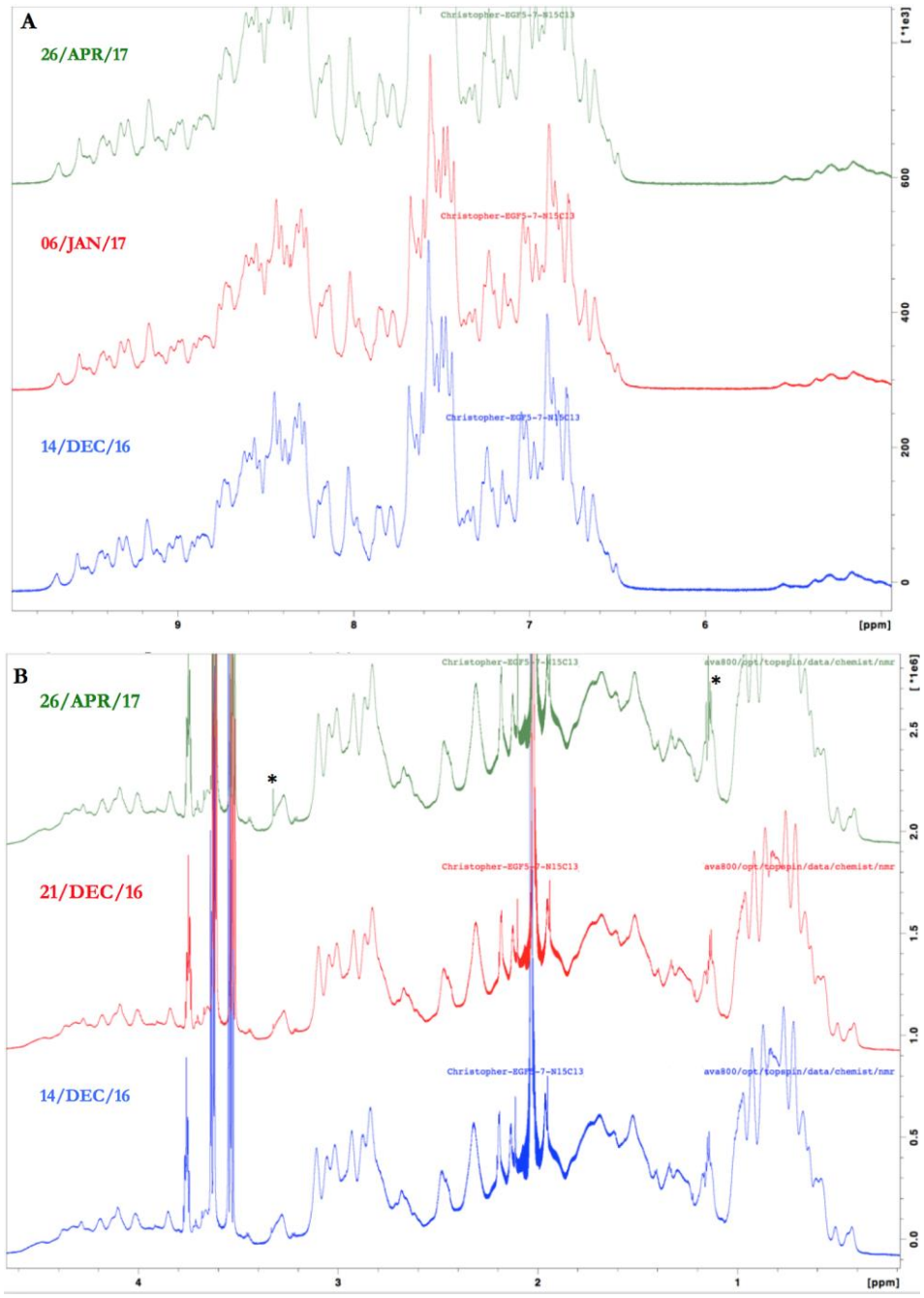


Figure 5.40. 1D NMR Monitoring of $^{15}N/^{13}C$-EGF5-7 Over Time: A. The amide region zoomed in from when the 25 °C experiments commenced (blue), after one week of continuous sample exposure to the 800 MHz spectrometer (red), and during the 40 °C experiments (green). B. The methyl region zoomed in. Asterisks' highlight anomalous peaks. The April measurements were taken under a week before commencing the relaxation experiments on the 600 MHz spectrometer (and immediately after finishing the 800 MHz spectrometer relaxation experiments). 24 scans carried out.

As an additional test of sample quality and robustness, a series of HSC spectra were also recorded (figure 5.41). No peak shifts or loss of signal over time were observed, and all peaks remained well dispersed. These observations indicate the protein remained natively folded over many weeks. This data is suggestive of a strong and stable protein, which is of benefit for follow-up studies in the future.

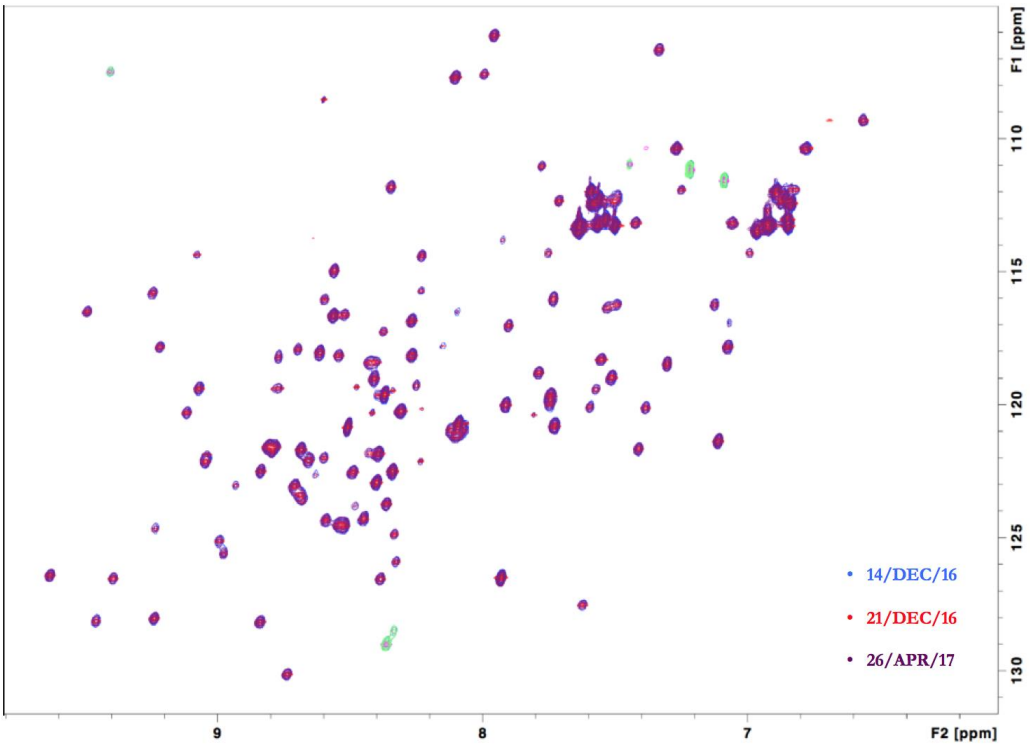


Figure 5.41. ^{15}N-HSQC Monitoring of ^{15}N/^{13}C-EGF5-7 Over Time: An overlay of each HSQC obtained from the dates displayed. Folded peaks appear as green and pink. Acquisition time was 30-minutes.

5.6 EGF-Rich PfRipr Protein Binding Assays to CyRPA

5.6.1 Methods: EGF-Rich PfRipr Protein Binding Assays to CyRPA

5.6.1.1 Gel-Filtration Based Binding Assays

Experiments were carried out at the University of Edinburgh by the author, and repeated at the Walter and Eliza Hall Institute by Dr. Wilson Wong using the PfRipr EGF-rich recombinant proteins produced in Edinburgh. Ripr EGF domains (10 μg for EGF1-2, 10.6 μg for EGF5-7; 15 μg for EGF7-10; 10 μg EGF3-10) were incubated with CyRPA (30 μg) at equal molar ratio in 1 × PBS at room temperature for one-hour. Samples were separated on a size exclusion column (Superose 6 10/300) in elution buffer (20 mM Tris, 150 mM NaCl, pH8) and peak fractions were analysed by reducing SDS-PAGE. The use of increased protein amounts was also attempted (for example, 40 μg of CyRPA and 80 μg of EGF5-7). The testing of EGF3-10 was not done in duplicate (only done at Edinburgh).

5.6.1.2 ELISA-Based Binding Assays

ELISA assays carried out using midisorp 96-well plates (Tomtec) and a SpectraMax M5 multimode plate reader at the EPPF. All proteins were prepared in PBS, blocking steps carried out using 5% w/v skimmed milk powder made up in SuperBlock (ThermoFisher). Washing steps and detection was done using methods described in section 3.2.2. Unless otherwise stated, all steps were carried out at room temperature. Wash steps were carried out between each step. Two methods were employed.

The first method involved coating the plates with 2 μg of CyRPA (donated by Dr. Wilson Wong) overnight at 4 °C, then blocking for one-hour. 4 μg of the EGF

proteins (EGFs 1-2, 3-6, 5-7, 7-10, 3-10) was titrated out in a 1:2 dilution series then added to the CyRPA-coated wells and incubated for 1.5-hours. A 1:4000 dilution of rabbit polyclonal anti-PfRipr antibodies was added (a 1:1000 solution of anti-EGF1-2 antibodies used for the EGF1-2 sample) and incubated for one-hour, followed by the addition of the HRP-conjugated antibody (1:2000) for another hour, then developed. Controls included CyRPA>Rh5>αRh5>detection (positive control), CyRPA>BSG>αBSG>detection (negative control), Rh5>BSG> αBSG>detection (positive control), each protein on their own detected with their cognate antibody, blank coating (only receiving antibodies and detection reagent), and blocking solution>each protein>protein detection.

The other way the ELISA was set up was to do as above but coat plates with 5 μg of CyRPA titrated in a 1:2 dilution across the wells, then add 2 μg of the secondary proteins.

5.6.1.3 NMR-Based Binding Assays

A sample of ^{15}N-EGF5-7 was prepared and aliquoted into four 0.5 ml Eppendorf sample tubes at a concentration of 50 μM in 20 mM MES pH 6.55, 75 mM NaCl, 10% v/v D_2O (this was the buffer matching the CyRPA sample). A 1D and 2D ^{15}N-HSQC experiment was carried out on the sample for the starting control (45-minute acquisition time at 25 °C), then 50 μl of 1, 5 and 10 molar excess of CyRPA was added to each of the three remaining sample tubes containing ^{15}N-EGF5-7, these were incubated at room temperature for ten-minutes, then analysed for chemical shift perturbations in HSQC spectra. Topspin software was used to assess the overlays of the resultant spectra.

5.6.2 Results: EGF-Rich PfRipr Protein Binding Assays to CyRPA

5.6.2.1 Gel-Filtration Based Binding Assays

None of the PfRipr EGF-rich proteins tested (EGFs 1-2, 5-7, 7-10 and 3-10) showed evidence of binding to CyRPA (full data not shown). An example of the negative results obtained is shown in figure 5.42, using EGF5-7 produced in SS *Pichia* (other proteins were produced in *E. coli* aside from EGF7-10). These results may be indicative that a non-EGF region of PfRipr is involved in binding to CyRPA. However, since EGF3-10 was the product of protein refolding and was never fully biophysically characterized, it cannot be ruled out that EGF3 or 4 is not involved in binding. Unfortunately, the refolded EGF3-6 material that was produced had been used up by the time of testing, and time did not permit further purification. Another alternative explanation is that the complex formed is relatively weak and unstable/transient. Due to this it was decided to pursue an assessment of binding using other methods such as ELISA and NMR where possible.

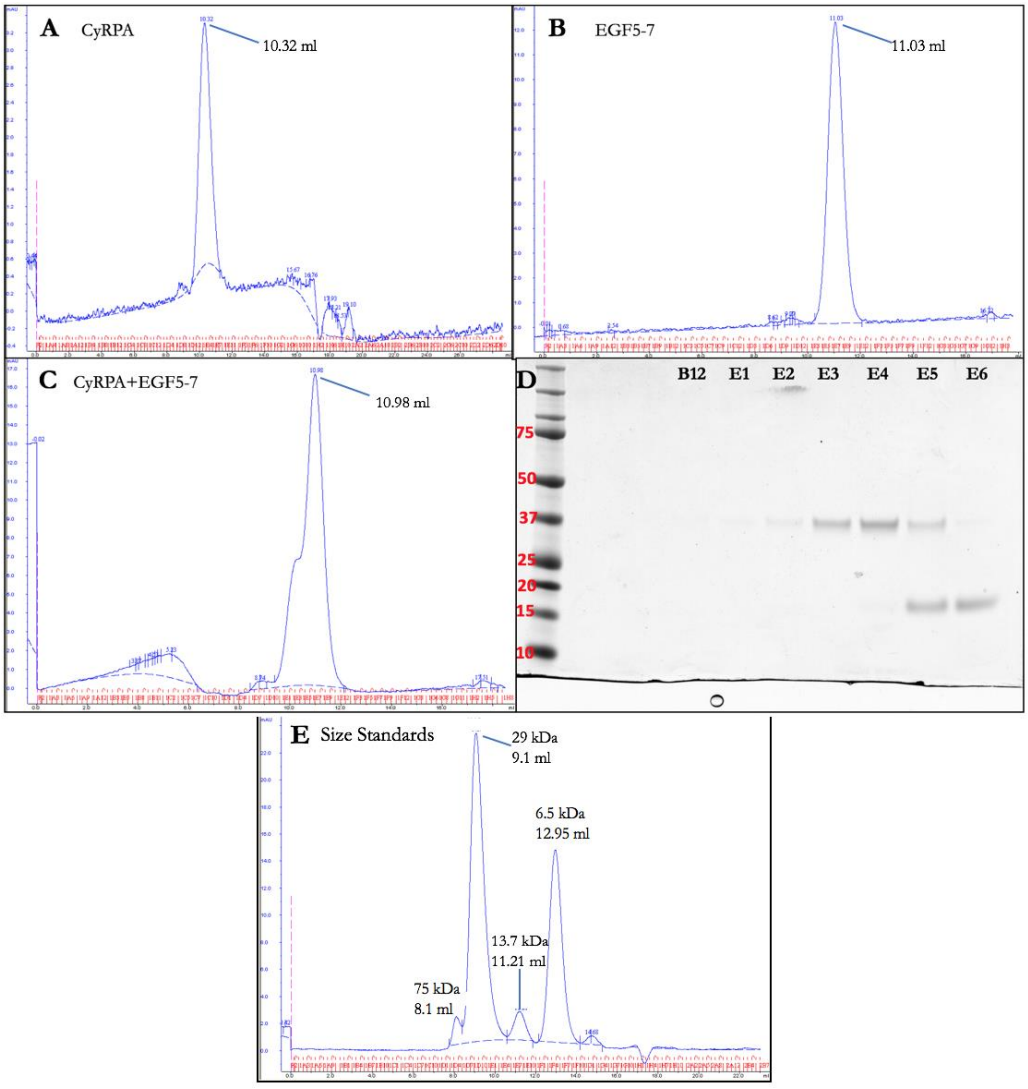


Figure 5.42. SEC-Based Binding Assay Between CyRPA and PfRipr EGF-Rich Proteins: A. CyRPA only. B. EGF5-7 only. C. CyRPA + EGF5-7. D. SDS-PAGE gel of various fractions taken from under the curve. E. Molecular weight size standards given in kDa and where they elute.

5.6.2.2 ELISA-Based Binding Assays

After numerous attempts, the initial ELISA set-up yielded positive results for the controls only (aside from EGF1-2 which was not detected); no experimental wells yielded any results as all appeared to show equal levels of positive signal come

through at the same strength (data not shown). This also occurred when the alternative set up of the ELISA was carried out (data not shown). This was suggestive that perhaps the protein concentration or antibody concentration was not optimal. However due to the focus on NMR-based studies, this was unable to be followed up.

5.6.2.3 NMR-Based Binding Assays

NMR can be a powerful tool for screening the weak interactions between biomacromolecules (Vinogradova & Qin, 2012). It was decided to utilize this approach as the SEC results implied that any interactions between CyRPA and EGF fragments are likely to be weak. Given the abundance of ^{15}N-EGF5-7 that had been produced, and there was only enough CyRPA to carry out the one experiment (ergo, no ^{15}N-EGF7-10 was able to be tested). However, no peak shift or intensity change was shown when comparing each CyRPA titration to the ^{15}N-EGF5-7-only sample (data not shown). This became conclusive evidence that the EGF5-7 region of PfRipr was not involved in forming either a tight or weak intermolecular complex with CyRPA.

5.7 Ca^{2+} Titration Study of PfRipr Proteins

5.7.1 Ca^{2+} Titration Study Methods

NMR experiments were performed on an AVANCE 800 MHz (Bruker) spectrometer using a 5-mm triple resonance cryoprobe, with experimental set-up performed by Juraj Bella. NMR data was acquired using Bruker TOPSPIN 1.2 and analysed using MestReNova software and CCPNMR Analysis. Advice on the design of the titration study was provided by Prof. Dusan Uhrin and Dr. Andrew Herbert. Five 0.5 ml aliquots of 40 μM ^{15}N-EGF5-7 were prepared in 20 mM MES

pH 6.55, 50 mM NaCl, 10% v/v D2O. 50 µl samples of $CaCl_2$ were prepared in the same buffer as the protein sample at concentrations of 2, 5, 10, and 20 mM. Data acquisition took place over 45-minutes.

5.7.2 Ca^{2+} Titration Study Results

5.7.2.1 Ca^{2+} Titration Study Results for ^{15}N-Labeled EGF5-7

Given the role that PfRipr appears to play in Ca^{2+}-signally during erythrocyte invasion, and the fact that EGF-like domains are well-known for binding divalent cations, an NMR-based titration study was carried out starting with adding a physiologically relevant concentration of Ca^{2+} (with respect to the levels found in the blood) and titrating up to a concentration where any ^{15}N-HSQC-based peak shifts became ablated. When 2 mM of the $CaCl_2$ solution was added to the protein, a peak shift became apparent, leading to a continuation of the titration up until 20 mM (figure 5.43). It is evident that EGF5-7 was binding to the Ca^{2+} as significant chemical shift perturbations were seen for certain residues (i.e. E776 and D821) but not all, and the shifts were occurring in different dimensions (the ^{1}H dimension on the X-axis of figure 5.43 for E776, and both the ^{1}H and ^{15}N dimension of the Y-axis for D821).

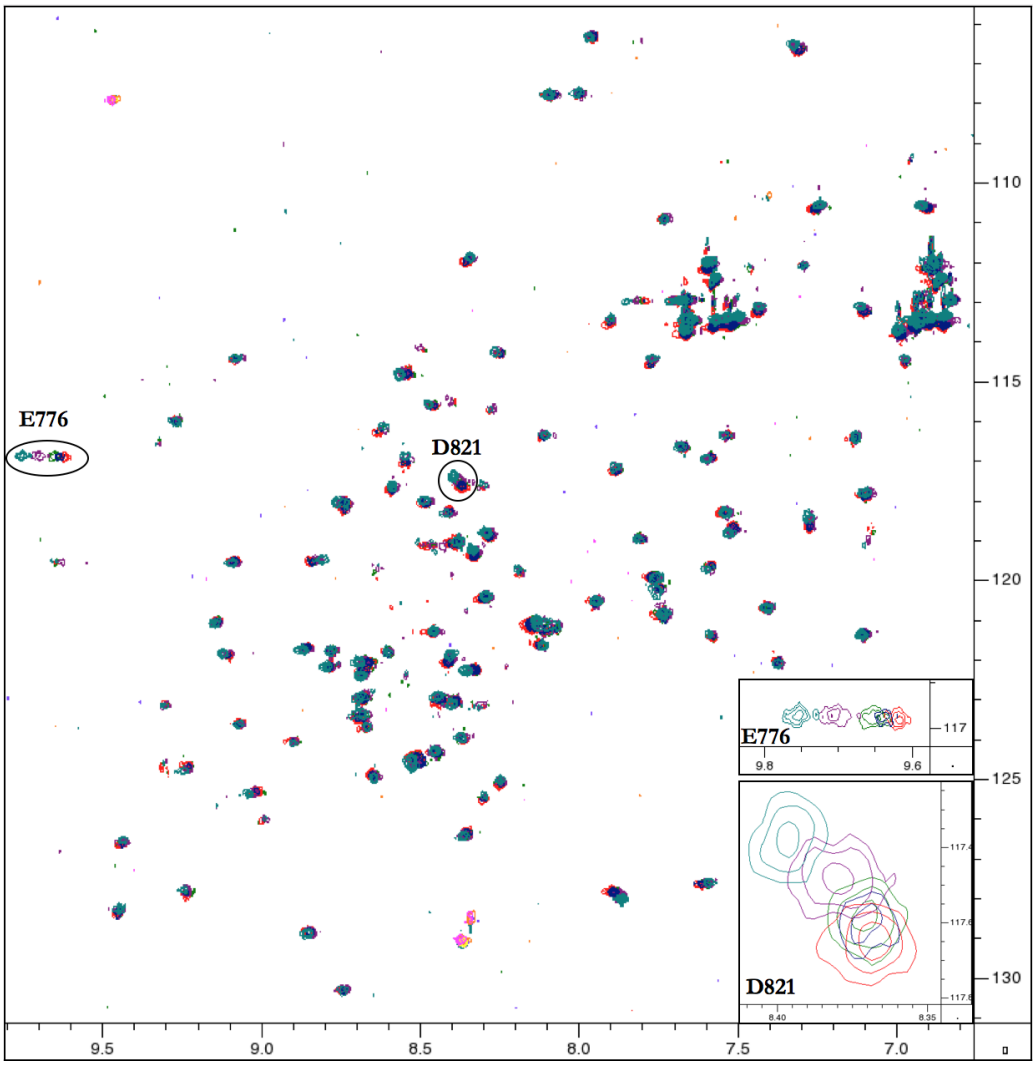


Figure 5.43. Ca²" Titration of ^{15}N-EGF5-7 via NMR: ^{15}N-HSQC of a Ca^{2+} titration into a solution of ^{15}N-EGF5-7. 0 mM $Cacl_2$ shown in turquoise, 2 mM in purple, 5 mM green, 10 mM blue, and 20 mM red. An example of two restudies that underwent a significant chemical shift perturbation during the titration are highlighted, with a zoom in for E776 and D821. Folded peaks shown in pink and orange.

To map and identify all residues that underwent a significant or minor chemical shift (significant defined as at least 0.1 ppm, and minor defined as at least 0.035 ppm; using a more conservative cut-off than that provided in (Bardelli, *et al.*, 2015)), a histogram plot was produced, as shown in figure 5.44. Six residues were identified has having at least undergone a minor chemical shift perturbation, one of which (N779) was only just included, and E776 displaying a major chemical

shift. The majority of residues identified were found in EGF7; D821, K847, N849 and E856 (the latter three all lie towards the C-terminus of the protein sequence).

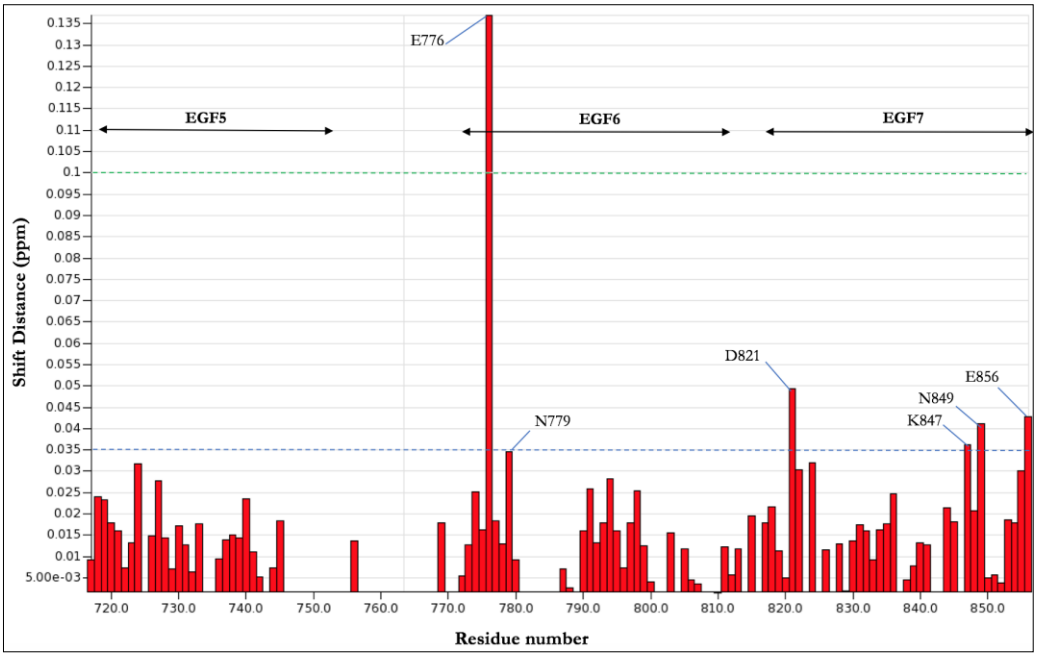


Figure 5.44. Histogram of ^{15}N-EGF5-7 Residue Shifts During Ca^{2+} Titration This plot demonstrates the chemical shift differences between cross peaks in spectra acquired in the presence of 0 mM and 20 mM of Ca Cl_2. The overall shift distance through both proton and nitrogen dimensions is given on the Y axis, which residues and where they are found in each EGF-like domain shown in the X-axis in in the plot, respectively. The blue dotted line shows the arbitrary cut-off of where a residue's shift perturbation was classed as "minor". The green dotted line is where the shift perturbation was considered "major". Residues that experienced at least a minor chemical shift perturbation are labelled.

Next, CCPNMR Analysis was used to determine a K_D value for each instance where a residue's chemical shift was measurably perturbed by Ca^{2+} binding. (figure 5.45). Overall, the fit of the data was relatively good, with E776 showing the best fit with the data. The residue showing the weakest affinity (that is, the highest K_D value) and the largest error margin was N779, which makes sense given that it showed the lowest shift distance in figure 4.54. The tightest binding affinity was with residue D821, with a K_D of 2.75 μM. All the residues identified displayed a

binding affinity around 100-times tighter than some other EGF-like domains in the literature that are shown to bind Ca^{2+} with high affinity (Handford, *et al.*, 1990).

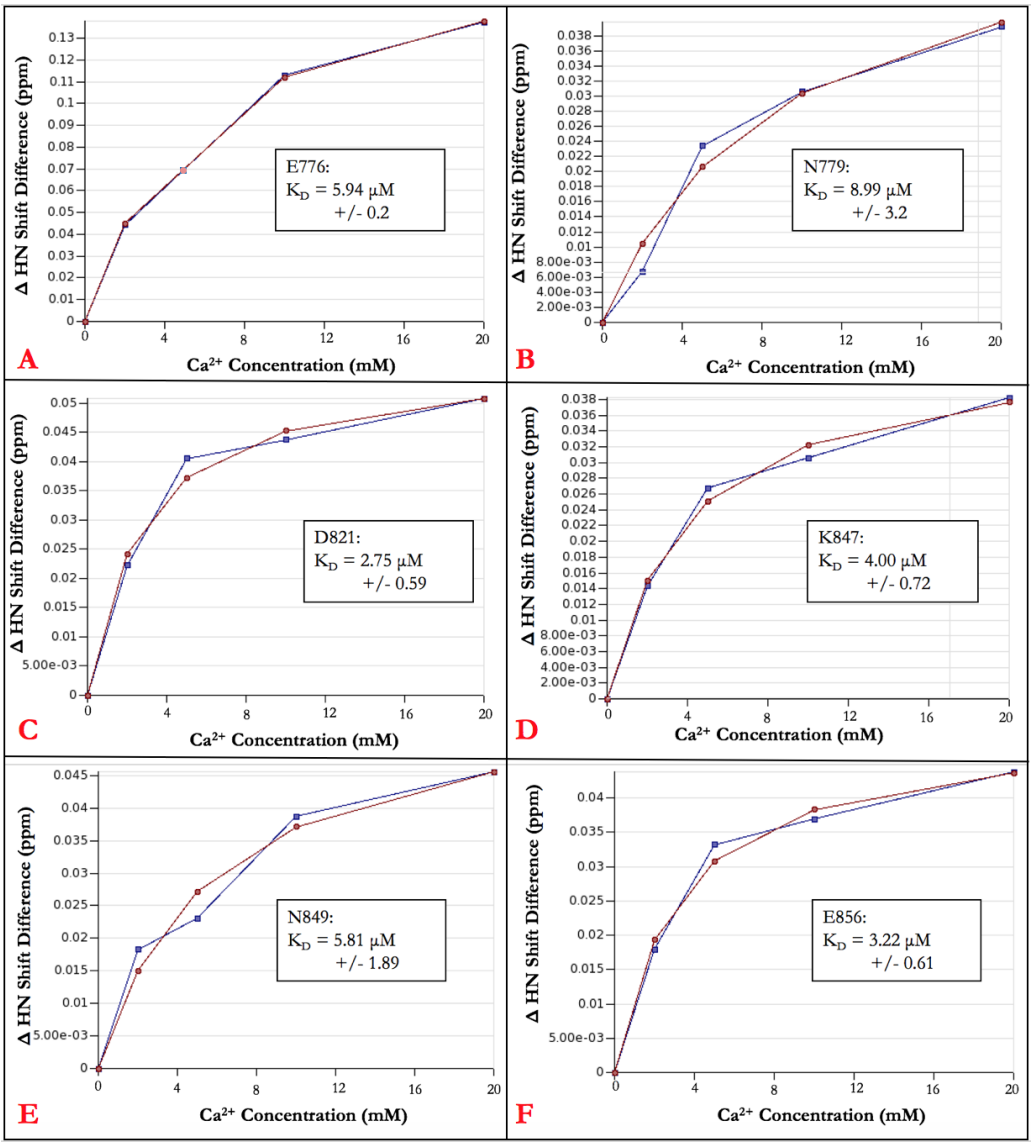


Figure 5.45. Binding Affinities for Each Ca^{2+}-Binding Residue of ^{15}N-EGF5-7: Each plot shows the fitting and the data points. Fitting functions for all used the equation A(B+xsqrt/((B+x)^2-4x)). A. The fitted graph for the shift distance of E776 as $CaCl_2$ was titrated into the solution. The insert shows the K_D value including the error margin. The value for the data point at 5 mM $CaCl_2$ was absent due to an unknown software issue. B. The fitted graph for the shift distance of N779 as $CaCl_2$ was titrated into the solution. C. The fitted graph for the shift distance of D821 as $CaCl_2$ was titrated into the solution. D. The fitted graph for the shift distance of K847 as $CaCl_2$ was titrated into the solution. E. The fitted graph for the shift distance of K849 as $CaCl_2$ was titrated into the solution. F. The fitted graph for the shift distance of E856 as $CaCl_2$ was titrated into the solution.

5.8 Mapping of Inhibitory mAbs to PfRipr

5.8.1 Methods for the Mapping of Inhibitory mAbs to PfRipr

The mAbs developed at the Walter and Eliza Hall Institute had been screened via a GIA using standard methods as described in Lim, *et al.*, 2015. In brief- Trophozoite stage parasites at 0.1% parasitaemia were grown in a 50 µl culture at 2% haematocrit in 96-well round bottom microtitre plates (Falcon) with doubling dilutions of antibody. After incubation for 96-hours each well was fixed at room temperature for 30-minutes with 50 µl of 0.25% glutaraldehyde (ProSciTech) diluted in PBS. Following centrifugation at 900 × g rpm for two-minutes, supernatants were discarded and trophozoite stage parasites were stained with 50 µl of 5 X SYBR Green (Invitrogen) diluted in PBS. The parasitemia of each well was determined by counting 50 000 cells by flow cytometry using a Cell Lab Quanta SC – MPL Flow Cytometer (Beckman Coulter). Growth was expressed as a percentage of the parasitaemia obtained using a non-immune rabbit IgG control. All samples were tested in triplicate.

Identification of mAbs that recognise a conformational or linear epitope was carried out via Western blotting (using standard methods as previously described) using R and NR FL-PfRipr (10 ng).

E. coli-produced EGF1-2 and SS. *Pichia*-produced EGFs 5-7 and 7-10 were provided to Jenny Thompson for analysis alongside FL-PfRipr (produced in a *Drosophila* expression system) to test reactivity via Western blotting using mAbs raised against EGF3-10, and rabbit polyclonal anti-FL-PfRipr antibodies. This was carried out in duplicate.

5.8.2 Results for the Mapping of Inhibitory mAbs to PfRipr

Following on from the development of mAbs raised from animals inoculated with EGF3-10, a GIA was carried out to assess which mAbs (if any) inhibited invasion by 3D7 *P. falciparum* parasites entering human erythrocytes (figure 5.46). The results of the GIA demonstrated two mAbs, 5G6 and 1G12, that led to ~75% less parasite survival at an IgG concentration of 2 mg/ml. The antibody concentration where 50 % of parasites had perished was ~0.45 mg/ml for 1G12 and ~ 0.6 mg/ml for 5G6.

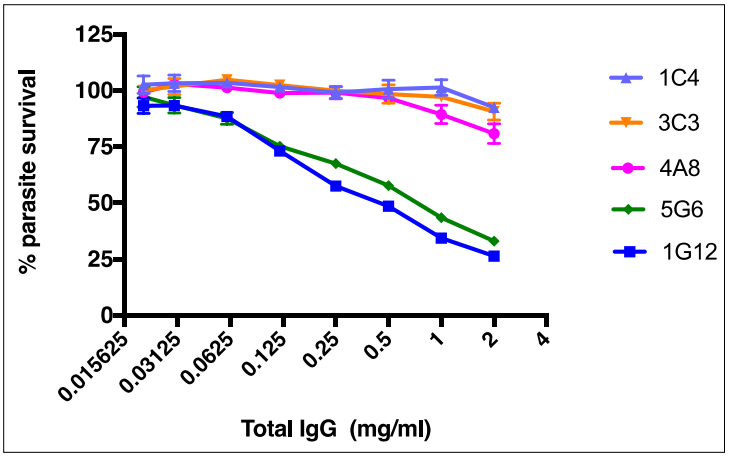


Figure 5.46. GIA Assay for Screening Anti-PfRipr mAbs: Parasite survival (Y-axis) is plotted against the total concentration of mAb IgG used (X-axis) for each mAb. Results are from the mean results from experiments carried out in triplicate, the error bars representing standard error.

Next, to test which mAbs recognised a conformational or linear epitope on PfRipr, a Western blot was carried out on R and NR FL-PfRipr. If antibodies bound only the NR form of the protein, they were judged to be recognising a conformational epitope. The results shown in figure 5.47 show that both inhibitory mAbs (5G6 and 1G12) identify a conformational epitope on PfRipr given the dearth of signal seen in R lanes. This is in contrast to mAbs 3C3 and 4A8, where clear bands are shown in both R and NR lanes, indicative of them binding to a linear epitope.

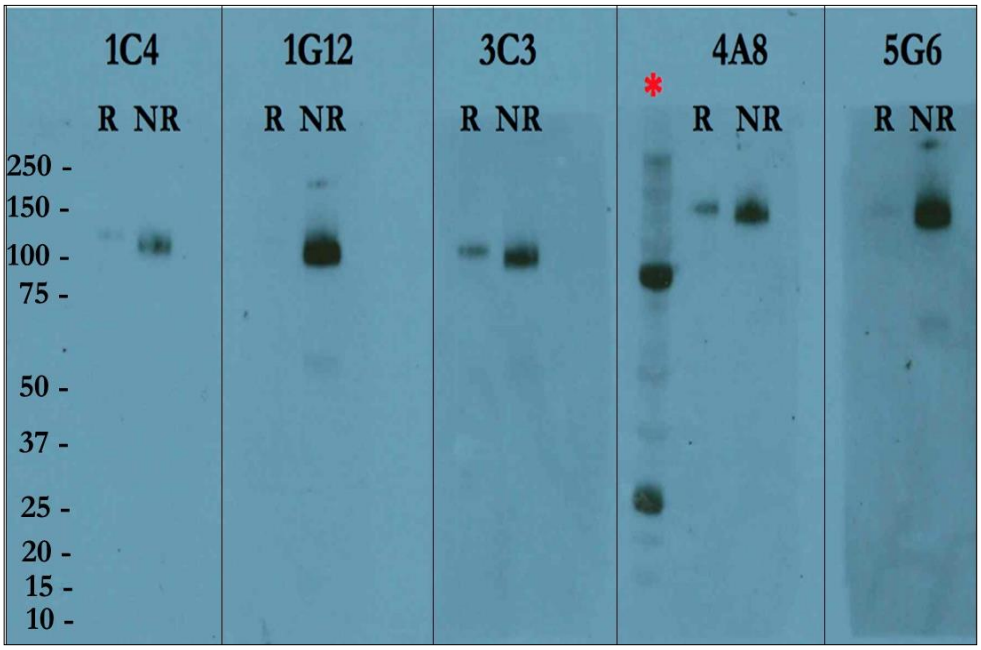


Figure 5.47. Identification of mAbs that Recognize a Conformational Epitope of PfRipr: Each mAb is listed atop the figure, with FL-PfRipr run on the gels under N and NR conditions. The lane denoted by a red asterisks (*) is a positive control for EGF3-10, detected by the polyclonal antibodies raised to it. Exposure time was 2.5-minutes. Molecular weight marker regions are given on the far left side.

Having established that 5G6 and 1G12 were inhibitory mAbs that bound to a conformational epitope of PfRipr, somewhere in the 3-10 region (indicating that there was a functional element somewhere here that the parasite relied on for survival in the blood stage of infection), it was important to narrow this down to either a non-EGF region, or if it did in fact bind to an EGF-containing region, which EGF the mAb bound to. For this the library of EGF-rich constructs were screened including EGF1-2 for a negative control (since the mAbs had been raised to the EGF3-10 region of PfRipr), FL-PfRipr as a positive control (and to serve as an identification of EGF3 or EGF4 if the mAbs did not recognise EGFs 5 – 10), EGF5-7 and EGF7-10 produced in SS. *Pichia*. Whilst there were no recombinant proteins of the non-EGF regions of PfRipr or the peptide that served as the linker between EGFs 5 and 6, it could be inferred as a possibility that the mAbs bound a

non-EGF region if they only bound FL-PfRipr and EGF5-7 (in addition to the possibility that the region being recognized was EGF5 or 6); if this were to be the result it would highlight the necessity of developing a refolding method for the non-EGF region-producing clones of *E. coli* described in section 3.2 of this thesis (or developing these constructs in SS. *Pichia*), and making the 17 amino acid peptide that comprised the long linker between EGF5 and 6.

The results in figure 5.48 show that only the inhibitory mAbs 5G12 and 5G2 bound to EGFs 5-7 and 7-10, with all other mAbs detecting only FL-PfRipr. The commonality between these proteins is the presence of the 7th EGF-like domain of PfRipr. These results suggest that the invasion-inhibiting mAbs bind to a conformational epitope present in EGF7. Furthermore, from the results of the GIA, it can be inferred that EGF7 is responsible for a key functional role during the process of invasion with the PAIN-complex. This should be regarded as a significant find, which is why a co-crystal screen has now been set up between these mAbs and EGF5-7 (with analysis of crystal wells on-going, data not shown).

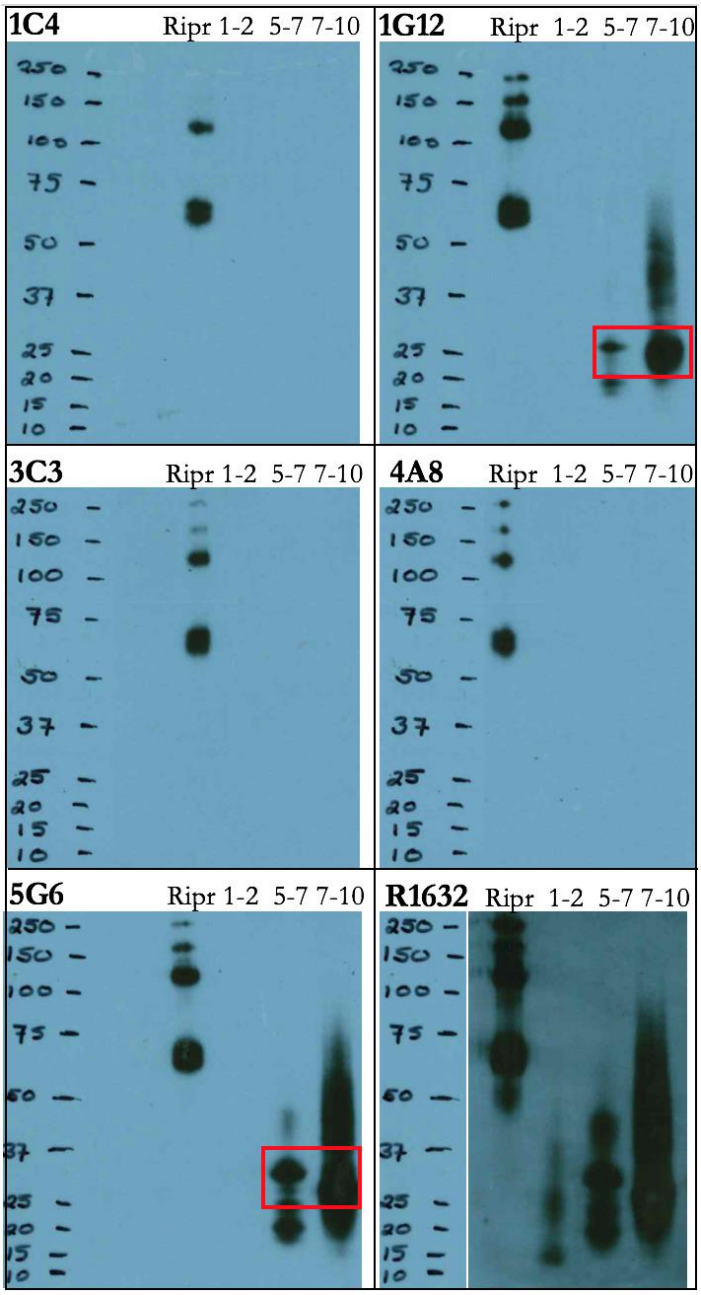


Figure 5.48. Inhibitory mAb Screening of PfRipr Proteins: The Western blots show each PfRipr protein screened for recognition by the anti-PfRipr (EGF3-10) mAbs. Here, "Ripr" denotes the FL-PfRipr protein. R1632 is a rabbit poly clonal anti FL-PfRipr antibody used as a positive control for detection of each protein. The red box highlights the EGF5-7 and EGF7-10 proteins detected by mAbs. As expected, no mAbs detected EGF1-2 (only the polyclonal raised to FL-PfRipr did). Exposure time for development was 2.5-minutes. These are the results of the second Western blot carried out, using less protein in an attempt to avoid over-loading the lanes.

5.9 Discussion and Conclusions

While it was unfortunate that FL-PfRipr was unable to be produced in the SS. *Pichia* system (along with EGFs 1-2, 3-5, 3-10, and the front half of PfRipr), two high-protein expressing clones were able to be developed; EGF5-7 and EGF7-10. Very likely it was merely a case of not obtaining any successful integrants with respect to the SS. *Pichia* transformed with the linearized plasmids (aside from the one Ripr21-1086/FL-PfRipr that was detected by Western blot in figure 5.2 A that unfortunately become contaminated and irretrievable). This could potentially be amended by scaling up the test expression from a handful of colonies to ~100.

However, the successful production of the malaria antigens comprising EGFs 5-7 and 7-10 in a Eukaryotic system should not be seen as overly trivial; competing malaria research groups using Eukaryotic systems for protein expression have had great difficulty with not only getting a malaria parasite protein produced, but produced in large quantities for in-depth biochemical and biophysical characterization (Dr. Julian Rayner, University of Cambridge, personal communication August 2017). Additionally, whilst generally recombinant proteins in these systems will be produced in their folded state, there is no guarantee of this to always be the case (depending on various factors such as intracellular stress during protein production cycles). In *P. pastoris*, low yields and proteolytic digestion of recombinant proteins have been reported to prevent the successful production of several proteins (reviewed in Ahmad, *et al*., 2014).

The results of this chapter overall show that regions of PfRipr (a notoriously difficult protein to work with as demonstrated by this thesis and the personal communications received from Prof. Alan Cowman and Dr. Julian Rayner) were successfully produced using the SS. *Pichia* system that allowed for 1 g of protein produced per L of supernatant (figure 5.1). Further, both EGF5-7 and 7-10 were not produced in a glycosylated state, a common problem for recombinant proteins produced in Eukaryotic systems. This in combination with the ability to utilize

large-scale fermenters present a significant cost-saving benefit for those with an interest in industrially-scalable systems of vaccine production. Additionally, the high yield of protein compensated for the loss of yield that often accompanies growth of cells in minimal media allowing facile uniform ^{15}N and ^{13}C isotopic labelling (figure 5.5 shows the high quantity of protein produced in minimal media).

As mentioned above, biophysical characterization of the proteins was of high importance. Once the correct mass had been confirmed, the evidence for correct protein folding needed to be gathered; the best evidence for this came from the NMR analysis of the 1D and the ^{15}N-HSQC spectra showing clear peak dispersion, lack of clustering of peaks in the HSQC (figure 5.23 A and C for EGF5-7, and figure 5.25 A and B for EGF7-10). Further confirmation of correct folding came from the assessment of free cysteine residues via the ESI-MS analysis of NEM and NEM + TCEP-treated EGF5-7. If the protein was not folded, or incorrectly folded then NEM would be able to react with and alkylate the free thiol on each cysteine. This was clearly not the case as shown in figure 5.29, where no increase in mass was detected.

The use of CD allowed for the probing of information about the secondary structure of the proteins that would help complement the NMR analysis, and provide some evidence as to the accuracy of the 3D modelling (via the homology-based modelling). Overall, the CD data in figure 5.18 suggested that EGF5-7 had a mostly unordered structure devoid of secondary structural elements, with the exception of β-strands that made up an estimated 35% of the protein. This approximately matched the SWISS-Model modelling data that gave an estimate of ~31% β-strand present in the protein (table 4), which was seen on the model that was processed in figure 5.28 A. Both EGF5-7 and 7-10 models showed elongated proteins comprised of EGF-like domains in tandem which in turn are composed of loops and β-sheets. Additionally, evidence of β-sheet for both proteins was seen

in 1D NMR spectra, where the ^{1}H peaks in the 5 – 5.5 ppm region are indicative of Hα atoms in a β-sheet.

One interesting pair of results were the comparison of the CD with that of the relaxation experiments including the heteronuclear NOE for EGF5-7. The CD of course suggested a highly unordered protein (usually suggestive of a flexible protein), meanwhile both sets of data (from the 600 and 800 MHz field strength NMR spectrometers) from the T_1, T_2 and heteronuclear NOE measurements in figures 5.36 – 5.38 showed strong evidence that the protein was in fact quite rigid (which was also observed in the SAXS data shown in figure 5.28 C and D), particularly in the EGF6 domain. This seemingly contradictory set of data may in fact be explained by the concentration of protein used in the experiments, with a significantly high concentration (~0.5 mM) used for the NMR experiments as compared to the much lower concentration used in the CD experiments. It may be that at higher concentrations, EGF5-7 undergoes a level of intermolecular self-association via EGF6. This would explain why the SEC analysis of the protein size did not show the presence of dimer when compared to the size standards in figure 5.42 E.

A follow up experiment should repeat the relaxation experiments with a protein at a lower concentration to look for evidence of self-association. Further, SEC-MALS (multi-angle light scattering) would be informative for this purpose (unfortunately due to a combination of equipment being unable to be used and a paucity of time, this was not able to be carried out). Additionally, in the future, relaxation experiments should be performed on EGF7-10 at both high and low concentrations to compare and contrast the results with that of EGF5-7 in order to evaluate the rigidity of the EGF-like domains at the very end of the C-terminus of the PfRipr protein. Finally, the relaxation data did indeed reveal that there were more flexible regions interspersed throughout EGF5-7, mostly in the flanking EGF-domains and the N- and C-termini. Various loop regions may also be stabilized by divalent cations; a follow up experiment would be to perform NMR

and CD analysis on samples with and without these ions, such as Ca^{2+} cations (via the use of a chelation agent).

The evidence for EGF5-7 being rigid conflicts with the original hypothesis that the reason all the crystallization attempts had failed was due to the protein being flexible. However, as mentioned above there was flexibility within the protein at various points, and these in combination with a partially self-associating protein would still potentially present thermodynamically unfavourable conditions for crystallography, particularly when the trials used proteins significantly over 0.5 mM in concentration (including ~1.5 mM). This would be particularly true if there was heterogeneous intermolecular self-association, since homologous oligomerization would actually promote crystal formation (W. Nicholson Price, *et al.*, 2009). Due to the discovery of mAbs that bind to EGF5-7 and 7-10 (as shown in figure 5.48), it will be worth screening crystal conditions using both antibodies and various protein concentrations, particularly lower concentrations.

Other findings from the biophysical characterization of the proteins described in this chapter include the evidence that the protein is relatively stable over time (as evidenced from the lack of spectra changes via 1D and 2D NMR analysis in figures 5.40 and 5.41 respectively), and also at high temperatures. The lack of denaturation at 40 – 70 °C as shown by the NMR temperature experiment in figures 5.19 – 5.21 was quite remarkable, although there was evidence that when kept at a higher temperature the protein would precipitate. However, after the sample conditions had been optimized, it was good knowing that EGF5-7 could happily tolerate the 25 - 40 °C temperatures, so that a full set of triple-resonance NMR experiments could be carried out at a variety of temperatures for the collection of data required for 3D structure elucidation of a $^{15}N/^{13}C$-EGF5-7 sample.

Regarding the $^{15}N/^{13}C$-EGF5-7 sample, it has been a great shame that the 3D structure has not been able to be solved. This was due to a paucity of resources leading to a lack of time towards the end of the research project. However, all the experimental data from several temperatures have been collected that would allow for this to be solved in the near future. Furthermore, high-quality SAXS data (figure 5.28 D) will help refine future structural models that come from completion of the 3D-NMR structure. Regarding the NMR analysis, the backbone amino acid assignment sits at 86% (as shown in figure 5.34) and will be improved once the side-chains become mapped out. Further, the backbone assignment had proven invaluable in the analysis of the Ca^{2+}-binding data, with figures 5.43 and 5.44 highlighting clear peak shifts during the titration of increasing concentrations of $CaCl_2$, and the residues involved. This is the first step taken to confirm that PfRipr may indeed utilize Ca^{2+} during the invasion process. Initial work in Volz, *et al.*, 2016 revealed using a calcium-sensitive fluorescence indicator, Fluo-4AM, that Ca^{2+} intensity was increased between the apical tip (where the PAIN-complex is located) of invading merozoites and the erythrocyte. This was followed up by demonstrating that PfRipr-deficient merozoites (developed with the lox-cre recombinase system) showed no measurable Ca^{2+}-spike when interacting with erythrocytes. This suggested of course a role for PfRipr in the calcium-signalling. Figure 5.45 reveals the affinity (via K_D measurements) of the key residues of EGF5-7 that bind calcium, which are tighter than EGF-like domains in the literature that are called "high affinity" binders of Ca^{2+}. Of course, the next steps will be to examine the other regions of PfRipr to assess their interaction with Ca^{2+} and to then develop parasite strains with selective amino acid mutations in these regions to be used in repeat experiments of the ones described in (Volz, *et al.*, 2016).

Of course, one of the aims of this research was to identify the region of PfRipr that binds to CyRPA. Unfortunately, EGFs 1-2, 5-7 and 7-10 all failed to show an interaction with CyRPA, and there was not enough EGF3-10 or 3-6 left to test. This does however narrow down the binding region to EGF3-4 or a non-EGF

region. Therefore, more EGF3-6 and 3-10 should be refolded and purified for binding studies via SEC plus microscale thermophoresis which can detect weaker interaction between proteins without the need for isotopic labelling or the addition of a chemical tag (for use with NMR and FRET, respectively). Furthermore, the non-EGF regions of PfRipr should be produced to carry out binding studies via the aforementioned techniques.

Initially, the identification of a pair of inhibitory mAbs that bind EGF7 could have been suggestive of this region being involved in the association with CyRPA; however, the negative data from the repeats of the SEC binding studies suggest this to be unlikely unless the interaction is very weak. More likely, the inhibitory mAbs are interfering with another functional aspect of the protein. Given that PfRipr is processed into two polypeptides that then self-associate (as shown in Chen, *et al.*, 2011) and form a complex with CyRPA, a leading hypothesis would be that the mAbs may be interfering with this intermolecular self-association. This could be tested by first screening the PfRipr EGF-rich protein constructs for binding amongst each other (via SEC), then applying each inhibitory mAb and re-doing the SEC-binding assay with SDS-PAGE and Western blot analysis of the eluted fractions. It should be mentioned that the fact that the 1G12 and 5G6 mAbs recognize a conformational epitope is of significance since epitopes that are recognized by neutralizing antibodies are structured 3D electrostatic landscapes (as discussed in Dormitzer, *et al.*, 2008). This is positively correlated with the development of high-titre neutralizing antibodies.

Since inhibitory mAbs have identified key immunologically relevant regions of CyRPA and PfRh5 including their epitope mapping (due to co-crystal structures reported in Favuzza, *et al.*, 2017, Chen, *et al.*, 2017 and Wright, *et al.*, 2014), and the EGF7 domain of PfRipr has now been identified as the region where inhibitory mAbs bind, it will be important to establish and map out the exact epitope. These mAbs have now been used for co-crystal trials with EGF5-7, which if sufficient resolution were to be obtained (assuming crystals can be obtained), would reveal

the exact conformational epitope within EGF7. However, in lieu of a crystal structure, the NMR data garnered from the backbone assignment of $^{15}N/^{13}C$-EGF5-7 can be used for epitope mapping studies. Even though the backbone assignment is incomplete for the whole of EGF5-7, the entirety of EGF7 has been mapped. Therefore, it will be imperative to perform a titration study on a ^{15}N-EGF5-7 sample using the mAbs, or preferably, either a Fab fragment (production being relatively easy via enzymatic digestion of a mAb) or a single chain variable fragment (more challenging and sometimes impossible to produce as they must be made recombinantly), and record ^{15}N-HSQC spectra to identify peak changes (as described in Bardelli, *et al.*, 2015). Alternatively, a ^{13}C-HSQC could be carried out on the $^{15}N/^{13}C$-EGF5-7 sample, which has the advantage of analysing the CH groups which outnumber NH groups in a protein. Of interest, both a B-cell epitope (BCE) and T-cell epitope (TCE) have been bioinformatically predicted within EGF7 (Kariuki, 2014). The predicted BCE consists of residues 823 – 852, whereas the TCE consists of residues 853 – 861 (with the last five residues comprising the linker and start of EGF8) for CD8+ T-cell epitopes, and residues 828 – 845 for CD4+ T-cell epitopes.

Finally, in the near future, the optimization of the disulfide mapping using a buffer suited to tryptic digest will yield the disulfide bond patterns found in EGF5-7. This will help inform the NMR-based structural calculations which may be highly inaccurate if one assumes the typical disulfide pattern of most EGFs of C1-C3, C2-C4, and C5-6. Further, the unusual long linker between EGFs 5 and 6 may not be forming a disulfide bond between each other, but could in fact be disulfide-bonded to cysteines in the EGF regions.

Chapter 6

Final Discussion, Conclusions and Future Directions

"*Let me tell you the secret that has led me to my goal. My strength lies solely in my tenacity*"
Louis Pasteur.

This thesis has described a complex project encompassing a variety of multidisciplinary fields (such as parasitology, biochemistry, immunology, biophysics, biophysical chemistry and structural biology) which changed directions as new evidence emerged. The project started with an examination of PfRh4 that, at the time of commencing the research, was deemed a viable vaccine candidate. It was considered important to elucidate the binding site for CR1 on PfRh4.9, solve the structure of PfRh4.18 and develop a set of inhibitory mAbs. The CR1-binding region remained elusive because each PfRh4.9 alanine-scan mutant bound with approximately the same affinity to CR1 as the native protein. This observation implied that the mAbs that had been used to map a candidate binding region comprising 58 residues within PfRh4.9, were not competing directly for binding with CR1 but were impeding CR1 access to a nearby binding site *via* steric hindrance, due to the size of the antibody molecule. The failure of this approach suggested structure determination as an alternative. As PfRh4.9 was too unstable for use in crystal trials the production of PfRh4.18 was undertaken, but this required a refolding protocol that required significant optimization.

At this point, more literature was appearing that encouraged a shift in focus towards the proteins of the PAIN-complex as potential top vaccine candidate against *P. falciparum* malaria. Of particular importance was the finding that IgG responses against PfRh5 and PfRipr were strongly associated with protective immunity (Richards*, et al.*, 2013). In fact, the median protective association for anti-PfRh5 and anti-PfRipr antibodies were significantly better than any other *P. falciparum* antigen that had reached phase-1 or phase-2 vaccination trials. Notably, whilst the Richards paper still listed PfRh4 among the top nine prospective vaccine candidates, there are key differences between it and the proteins of the PAIN-complex. Crucially, the PAIN-complex contains PfRh5 that was not amenable to genetic knock-out (Baum, *et al.*, 2009). Furthermore, work in the Cowman lab at WEHI was, at this time, showing that all proteins in the PAIN-Complex are essential; this work was published several years later (Volz*, et al.*, 2016). Taken together, the essential nature of the PAIN-complex proteins, their ability to induce

strain-transcending inhibitory antibodies, the clinical relevance of the findings by Richards, *et al.*, 2013 and the novel idea from Richards that there was a trend between newer malaria antigens discovered and their increased importance as vaccine candidates (superseding competing antigens that had been under development for longer) suggested the legitimacy of the PAIN-complex proteins as the best vaccine targets.

The final piece of indirect evidence that convinced the author the association stated by Richards was not merely a *post hoc ergo propter hoc* fallacy was the observation that, despite PfRipr being identified as a lead prospective vaccine candidate, no publications focused on it had come out since the original paper by Chen *et al.* in 2011, which was a space of more than 3 years by the time this doctoral research focused away from PfRh4 towards PfRipr. This suggested that the protein was perhaps quite difficult to work, which was shown to indeed be the case. Most protein biochemists would likely agree that PfRipr was not an ideal protein to try to work on since it had both a large proportion of cysteines and an odd amount (87), with most cysteines predicted to be involved in disulfide bonding (making correct folding a challenge in any recombinant system). Additionally, PfRipr is relatively large at ~123 kDa, and is enzymatically processed into two self-associating polypeptides, meaning that trying to express the full-length protein may not work without the presence of the enzyme that led to the cleavage event, and this enzyme is still unknown. It was therefore decided to focus more exclusively on the PfRipr protein. However, with that said, PfRh4 remains a biologically fascinating protein, and should form part of further studies aimed at generating a full structural library of *P. falciparum* invasion ligands.

To tackle the proverbial beast that was PfRipr, it was decided to chop the protein up into potentially more manageable components. Truncations comprised of the EGF-like domains were deemed an attractive starting place since EGF-like domains are widely known for their involvement in protein-protein interactions and being tractable to recombinant production and structure determination. This

led to the production of a library of protein fragments that were produced in both *E. coli* and *P. pastoris* expression systems. This library now contains EGF1-2 (*E. coli*), EGF3-6 (*E. coli*), EGF5-6 (*E. coli*), EGF5-7 (*P. pastoris*), EGF7-8 (*E. coli*), EGF9-10 (*E. coli*), EGF7-10 ((*E. coli* and *P. pastoris*) and EGF3-10 (*E. coli*). The examples of EGFs 5-6, 3-6, 5-7 and 7-10 are the first PfRipr fragments to be produced in relatively large quantities and biophysically characterized. Furthermore, the results in figure 5.48 show for the first time, the use of functionally active inhibitory mAbs (that allay parasite survival *in vitro*) mapping a specific region within PfRipr that is inferred to be of significant functional importance for this protein in the PAIN-complex (based on the deleterious effect these mAbs had on parasite survival in the GIA shown in figure 5.46). From a pure vaccinology perspective, the fact that EGF5-7 remains stable over a long period of time is a bonus. Whilst the protein proved stable at 40 °C for almost a month, future work should test the stability at this temperature over a series of several months.

Whilst in its own right, the aforementioned work is of significance, the next step for EGF5-7 will be to combine the biophysical characterization described in this thesis (CD, NMR, mass spectrometry, NEM alkylation-based disulfide analysis, T_1/T_2 relaxation measurements and heteronuclear NOE data), the Ca^{2+}-binding evidence, backbone assignment of the HSQC and mAb data with a completed 3D NMR structure (attainable given the quality of the NOESY spectra, data not shown, personal communication with Dr. Brian Smith), mapping the 3D structure onto the cryo-electron microscopy derived PAIN-complex structure (work in progress from the Cowman lab), NMR-based epitope mapping, recognition of EGF5-7 by human immune sera (from a pool of individuals with non-sterilizing immunity to *P. falciparum*, and complete disulfide mapping. Recently, SAXS data have been collected on the EGF5-7 sample, which should help with the structure determination and in particular defining spatial arrangements of the suspected EGF-like domains. Additionally, if possible, a co-crystal structure of EGF5-7 (or

EGF7 alone) with either 1G12 or the 5G6 mAb will be ideal for confirmation on the atomic resolution detail of the interaction between epitope and paratope. These studies would guide future structural vaccinology work and the design of experimental vaccines against PfRipr.

Looking further ahead, an ideal experimental vaccine for testing on the blood stage of infection should include an in-tandem or combination design of EGF7 with CyRPA (an individual domain may be problematic given that the interface between the mAb and the protein spans several regions (Chen, *et al.*, 2017)), and the ~45 kDa PfRh5 truncation (potentially using only the second and third α-helices due to their epitope interacting with the inhibitory mAbs QA5 and 9AD4 as described in Wright, *et al.*, 2014). These could further be coupled with CSP and an antigen that targets the sexual cycle of the parasite in order to test a multi-life cycle stage, multi-valent vaccine based on both immunological evidence and rational, structural vaccinology principles (as reviewed in Dormitzer, *et al.*, 2008)

Malaria has proven to be a highly-complicated foe, one that has led to an enormous amount of resources spent on trying to fight it on various fronts (i.e. vector control, drugs, vaccines and diagnostics). If it is to be cut like the Gordian-knot (a reference to the apocryphal tale of Alexander the Great and his approach to solving a complicated problem; of note, Alexander is said to have possibly died from malaria, ironic given the metaphor presented here (Cunha, 2004)), it will likely require a precisely targeted attack. This will only be attainable with continued financial and intellectual contribution, directed cross-institutional collaboration, and a willingness to drop so-called "pet proteins" when they are no longer ideal as a vaccine candidate.

Bibliography

Adachi, K. *et al.*, 2014. Economics of Malaria Prevention in US Travelers to West Africa. *Clinical Infectious Diseases,* September, 58(1), pp. 11-21.

Adams, J. H., Blair, P. L., Kaneko, O. & Peterson, D. S., 2001. An expanding ebl family of *Plasmodium falciparum*. *Trends in Parasitology,* June, 17(6), pp. 297-299.

Adams, J. H., *et al.*, 1992. A family of erythrocyte binding proteins of malaria parasites. *Proceedings of the National Academy of Sciences,* 89(15), pp. 7085-7089.

Alonso, P., 2017. Malaria Prevention Works: Let's close the gap. *World Health Organisation,* April.p. 1.

Aniweh, Y. *et al.*, 2017. *P. falciparum* RH5-Basigin interaction induces changes in the cytoskeleton of the host RBC. *Cellular Microbiology,* May, 19(9).

Auterhoff, H., 1967. Nobel Lectures, Physiology or Medicine 1901-1921. *Elsevier Publishing Company.*

Bargieri, D.Y., *et al.*, 2016. *Plasmodium* Merozoite TRAP Family Protein Is Essential for Vacuole Membrane Disruption and Gamete Egress from Erythrocytes. *Cell Host & Microbe*, November, 20(5), pp. 618-630.

Bathurst, I. *et al.*, 1993. An Experimental Vaccine Cocktail for *Plasmodium-Falciparum* Malaria. *Vaccine,* 11(4), pp. 449-456.

Baum, J. *et al.*, 2009. Reticulocyte-binding protein homologue 5 - an essential adhesin involved in invasion of human erythrocytes by *Plasmodium falciparum*. *International Journal of Parasitology,* February, 39(3), pp. 371-380.

Baum, J., Pinder, M. & Conway, *et al.*, 2003. Erythrocyte invasion phenotypes of *Plasmodium falciparum* in The Gambia. *Infection and Immunity,* 71(4), pp. 1856-1863.

Baum, J., Richard, D. & Healer, J., 2006. A conserved molecular motor drives cell invasion and gliding motility across malaria life cycle stages and other apicomplexan parasites. *Journal of Biological Chemistry,* February, 281(8), pp. 5197-5208.

Beck, L. R. *et al.*, 1994. Remote Sensing as a Landscape Epidemiologic Tool to Identify Villages at High Risk for Malaria Transmission. *The American Journal of Tropical Medicine and Hygiene,* September, 51(3), pp. 271-280.

Beeson, J. G. *et al.*, 2016. Merozoite surface proteins in red blood cell invasion, immunity and vaccines against malaria. *FEMS Microbiology Reviews,* February, 40(3), pp. 343-372.

Bergmann-Leitner, E. S. & Leitner, W. W., 2013. Improving DNA vaccines against malaria: could immunization by gene gun be the answer?. *Therapeutic Delivery,* July, 4(7), pp. 767-770.

Besteiro, S., Dubremetz, J. & and Lebrun, M., 2011. The moving junction of apicomplexan parasites: a key structure for invasion. *Cellular Microbiology,* June 13(6), pp. 797–805.

Biau, D.J., 2011. In Brief: Standard Deviation and Standard Error. *Clin Orthop Relat Res*, September, 469(9), pp. 2661-2664.

Bill, R.M., 2014. Playing catch-up with *Escherichia coli*: using yeast to increase success rates in recombinant protein production experiments. *Frontiers in Microbiology*, March, 5(85).

Boddey, J. A. & Cowman, A. F., 2013. *Plasmodium* nesting: remaking the erythrocyte from the inside out. *Annual Reviews Microbiology,* June, 67, pp. 243–269.

Boyle, M. J. *et al.*, 2010. Isolation of viable *Plasmodium falciparum* merozoites to define erythrocyte invasion events and advance vaccine and drug development.

Proceedings of the National Academy of Sciences of the United States of America , 107(32), pp. 14378-14383.

Bullen, H. *et al.,* 2012. Biosynthesis, Localization, and Macromolecular Arrangement of the *Plasmodium falciparum* Translocon of Exported Proteins (PTEX). *Journal of Biological Chemistry,* 287(11), pp. 7871-7884.

Bushell, E. *et al*., 2017. Functional Profiling of a *Plasmodium* Genome Reveals an Abundance of Essential Genes. *Cell,* July, 170(2), pp. 260-272.

Bustamante, L. Y. *et al.,* 2013. A full-length recombinant *Plasmodium falciparum* PfRH5 protein induces inhibitory antibodies that are effective across common PfRH5 genetic variants. *Vaccine,* 31(2), pp. 373-379.

Caminade, C. *et al*., 2014. Impact of climate change on global malaria distribution. *Proceedings of the National Academy of Sciences of the United States of America ,* January, 111(9), pp. 3286-3291.

Chen, L. *et al*., 2011. An EGF-like Protein Forms a Complex with PfRh5 and Is Required for Invasion of Human Erythrocytes by *Plasmodium falciparum*. *PLoS Pathogens,* September, 7(9).

Chen, L. *et al*., 2014. Crystal structure of PfRh5, an essential *P. falciparum* ligand for invasion of human erythrocytes. *eLife,* October. 3.

Chen, L. *et al*., 2017. Structural basis for inhibition of erythrocyte invasion by antibodies to *Plasmodium falciparum* protein CyRPA. *eLife,* February, 6.

Chiu, C. Y. H. *et al*., 2014. Association of antibodies to *Plasmodium falciparum* reticulocyte binding protein homolog 5 with protection from clinical malaria. *Frontiers in Microbiology,* June, 5(314).

Clyde, D. F., McCarthy, V. C., Miller, R. M. & Woodward, W. E., 1975. Immunization of Man against Falciparum and Vivax Malaria by Use of Attenuated Sporozoites. *The American Journal of Tropical Medicine and Hygiene ,* May, 24(3), pp. 397-401.

Cockburn, I. A. *et al.*, 2004. A human complement receptor 1 polymorphism that reduces *Plasmodium falciparum* rosetting confers protection against severe malaria. *Proceedings of the National Academy of Sciences of the United States of America ,* January, 101(1), pp. 272-277.

Collins, F.S., and Tabak, L.A., 2014. NIH plans to enhance reproducibility. *Nature*, January, 505(7485), pp. 612-613.

Coppi, A. *et al.*, 2007. Heparan sulfate proteoglycans provide a signal to *Plasmodium* sporozoites to stop migrating and productively invade host cells. *Cell Host Microbe,* November, 2(5), pp. 316–327.

Cowman, A. F., Healer, J., Marapana, D. & Marsh, K., 2016. Malaria: Biology and Disease. *Cell,* October, 167(3), pp. 610-624.

Cox, F., 1970. Protective immunity between malaria parasites and piroplasms in mice. *Bulletin of the World Health Organisation,* 43, pp. 325-336.

Crabb, B. S., *et al.*, 2006. Invasion of Red Blood Cells by Malaria Parasites. *Cell,* February, 124(4), pp. 755-766.

Crabb, B. S. *et al.*, 1997. Targeted Gene Disruption Shows That Knobs Enable Malaria-Infected Red Cells to Cytoadhere under Physiological Shear Stress. *Cell,* April, 89(2), pp. 287-296.

Cranston, H. *et al.*, 1984. *Plasmodium falciparum* maturation abolishes physiologic red cell deformability. *Science,* January, 223(4634), pp. 400-403.

Croft, A. M., Winfield, C. R., Horsfall, M. & Quarrell, M. A., 2005. Direct health costs of occupationally acquired malaria in a military population in Europe. *Occupational Medicine,* March, 55(2), pp. 128-130.

Crompton, P. D., Pierce, S. K. & Miller, L. H., 2010. Advances and challenges in malaria vaccine development. *Journal of Clinical Investigation,* December, 120(12), pp. 4168-4178.

Crosnier, C. *et al*., 2011. Basigin is a receptor essential for erythrocyte invasion by *Plasmodium falciparum*. *Nature,* December, 480(7378), pp. 534–537.

Dame, J. *et al*., 1984. Structure of the gene encoding the immunodominant surface antigen on the sporozoite of the human malaria parasite *Plasmodium falciparum*. *Science,* August, 225(4662), pp. 593-599.

Das, S. *et al*., 2015. Processing of *Plasmodium falciparum* Merozoite Surface Protein MSP1 Activates a Spectrin-Binding Function Enabling Parasite Egress from RBCs. *Cell Host and Microbe,* October, 18(4), pp. 433-444.

Doi, Y. *et al*., 2010. ADF2 is required for transformation of the ookinete and sporozoite in malaria parasite development. *Biochemical and biophysical research communications,* June, 397(4), pp. 668-672.

Dolan, S. A., Miller, L. H. & Wellems, T. E., 1990. Evidence for a switching mechanism in the invasion of erythrocytes by *Plasmodium falciparum*. *Journal of Clinical Investigation,* 86(2), p. 618.

Douglas, A. D. *et al*., 2014. Neutralization of *Plasmodium falciparum* merozoites by antibodies against PfRH5. *Journal of Immunology,* January. 192(1), pp. 245-258.

Douglas, A. D. *et al*., 2011. The blood-stage malaria antigen PfRH5 is susceptible to vaccine-inducible cross-strain neutralizing antibody. *Nature Communications* 2, pp. 601.

Dreyer, A. *et al.*, 2012. Passive immunoprotection of *Plasmodium falciparum*-infected mice designates the CyRPA as candidate malaria vaccine antigen.. *Journal of Immunology,* 188(12), pp. 6225-6237.

Ellis, J. *et al.*, 1983. Cloning and expression in *E. coli* of the malarial sporozoite surface antigen gene from *Plasmodium knowlesi*. *Nature,* April, 302(5908), pp. 536-538.

Elsworth, B. *et al.*, 2016. Proteomic analysis reveals novel proteins associated with the *Plasmodium* protein exporter PTEX and a loss of complex stability upon truncation of the core PTEX component, PTEX150. *Cellular Microbiology,* 18(11), pp. 1551-1569.

Favuzza, P. *et al.*, 2012. Cysteine-Rich Protective Antigen (CyRPA) as promising blood-stage candidate protein for inclusion in a malaria subunit vaccine. *Malaria Journal,* October, 11(Supplement 1), p. 30.

Favuzza, P. *et al.*, 2017. Structure of the malaria vaccine candidate antigen CyRPA and its complex with a parasite invasion inhibitory antibody. *eLife,* February, Volume 6.

Felgner, P. *et al.*, 2013. Pre-erythrocytic antibody profiles induced by controlled human malaria infections in healthy volunteers under chloroquine prophylaxis. *Scientific Reports,* December 3(3549).

Forni, D., Pontremoli, C., Cagliani, R., Pozzoli, U., Clerici, M., and Sironi, M., 2015. Positive selection underlies the species-specific binding of *Plasmodium falciparum* RH5 to human basigin, *Molecular Ecology*, September, 24(18), pp. 4711-4722.

Franke, D., Jefferies, C.M., and Svergun, D.I., 2015. Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. *Nature Methods*, July, 12, pp. 419-422.

Galassi, F. M. *et al.*, 2016. The sudden death of Alaric I (c. 370–410 AD), the vanquisher of Rome: A tale of malaria and lacking immunity. June, 31, pp. 84-87.

Galaway, F. *et al.,* 2017. P113 is a merozoite surface protein that binds the N terminus of *Plasmodium falciparum* Rh5. *Nature Communications,* February 8(14333), p. 14333.

Gallup, J. L. & Sachs, J. D., 2001. The Intolerable Burden of Malaria: A New Look at the Numbers- The Economic Burden of Malaria. *American Journal of Tropical Medicine and Hygiene,* January, 64(1).

Gao, X., Gunalan, K., Yap, S. S. & Preiser, P. R., 2013. Triggers of key calcium signals during erythrocyte invasion by *Plasmodium falciparum*. *Nature Communications,* 4(2862).

Gilberger, T. *et al.*, 2003. A novel erythrocyte binding antigen-175 paralogue from *Plasmodium falciparum* defines a new trypsin-resistant receptor on human erythrocytes. *Journal of Biological Chemistry,* April, 278(16), pp. 14480-14486.

Gilson, P. *et al.*, 2008. Identification and stoichiometry of glycosylphosphatidylinositol-anchored membrane proteins of the human malaria parasite *Plasmodium falciparum*. *Molecular Cell Proteomics,* July, 5(7), pp. 1286-1299.

Godfrey, W. & Haggarty-Weir, C. N., 2017. Popper's Formulation of Scientific Knowledge: A shift from an inductivist account. *The Philosopher,* June.

Gordon, D. *et al.*, 1995. Safety, immunogenicity, and efficacy of a recombinantly produced *Plasmodium falciparum* circumsporozoite protein-hepatitis B surface antigen subunit vaccine. *Journal of Infectious Diseases,* June, 171(6), pp. 1576-1785.

Greenwood, B. & Duombo, O. K., 2016. Implementation of the malaria candidate vaccine RTS,S/AS01. *The Lancet,* January, 387(10016), pp. 318-319.

Griffin, J. T., Ferguson, N. M. & Ghani, A. C., 2014. Estimates of the changing age-burden of *Plasmodium falciparum* malaria disease in sub-Saharan Africa. *Nature Communications,* February, 5(3136).

Guttery, D., Holder, A. A. & Tewari, R., 2012. Sexual development in *Plasmodium*: Lessons from functional analyses. *PLoS Pathogens,* January, 8(1).

Haggarty-Weir, C. N., Galloway, R. & Godfrey, W., 2017. Methods for the Refolding of Disulfide-Rich Proteins. *BioArxiv,* June.

Hayton, K. *et al.*, 2008. Erythrocyte binding protein PfRH5 polymorphisms determine species-specific pathways of *Plasmodium falciparum* invasion. *Cell Host and Microbe,* July, 4(1), pp. 40-51.

Hempelmann, E. & Krafts, K., 2013. Bad air, amulets and mosquitoes: 2,000 years of changing perspectives on malaria. *Malaria Journal,* 12(232).

Herrera, R. *et al.*, 2015. Reversible conformational change in the *Plasmodium falciparum* circumsporozoite protein masks its adhesion domains. *Infection and Immunity,* October, 83(10), pp. 3771–3780.

Hill, A. V., 2011. Vaccines Against Malaria. *Philosophical Transactions of the Royal Society B,* September.

Holder, A. A., 1994. Proteins on the surface of the malaria parasite and cell invasion. *Parasitology,* 108, pp. S5-18.

Ito, D., Schureck, M. & Desai, S., 2017. An essential dual-function complex mediates erythrocyte invasion and channel-mediated nutrient uptake in malaria parasites. *eLife,* February, 6.

Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heynecker, H.L., Bolivar, F., *et al.*, 1977. Expression in *Eschericia coli* of a chemically synthesized gene for the hormone somatostatin. *Science*, December, 198(4321), pp. 1056-63.

Jayner, J. C. *et al.*, 2001. A *Plasmodium falciparum* homologue of *Plasmodium vivax* reticulocyte binding protein (PvRBP1) defines a trypsin-resistant erythrocyte invasion pathway.. *Journal of Experimental Medicine,* December, 194(11), pp. 1571-1581.

Jowett, M. & Miller, N. J., 2005. The financial burden of malaria in Tanzania: implications for future government policy. *The International Journal of Health Planning and Management,* January, 20(1), pp. 67-84.

Joy, D. A. *et al.*, 2003. Early Origin and Recent Expansion of *Plasmodium falciparum*. *Science,* April, 300(5617), pp. 318-321.

Karunamoorthi, K., 2014. The counterfeit anti-malarial is a crime against humanity: a systematic review of the scientific evidence. *Malaria Journal,* June, 13(209).

Kaushansky, A. *et al.*, 2015. Malaria parasites target the hepatocyte receptor EphA2 for successful host infection. *Science,* November, 350(6264), p. 1089–1092.

Kelly, S.M., Jess, T.J., and Price, N.C., 2005. How to study proteins by circular dichroism. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, August, 1751(2), pp. 119-139.

Kester, K. E. *et al.*, 2009. Randomized, Double-Blind, Phase 2a Trial of *Falciparum* Malaria Vaccines RTS,S/AS01B and RTS,S/AS02A in Malaria-Naive Adults: Safety, Efficacy, and Immunologic Associates of Protection. *The Journal of Infectious Diseases,* August, 200(3), pp. 337-346.

Koch, M., *et al*., 2017. *Plasmodium falciparum* erythrocyte-binding antigen 175 triggers a biophysical change in the red blood cell that facilitates invasion. *PNAS*, April, 114(16), pp. 4225-4230.

Kocovski, N., Godfrey, W., Elkington, D. & Weir, C., 2015. Reviewing anti-malarial usage and resistance patterns and its effects on world health organisation programs. *Russian Journal of Parasitology,* 3, pp. 65-74.

Koning-Ward, T. F. d. *et al*., 2009. A newly discovered protein export machine in malaria parasites. *Nature,* April, Volume 459, pp. 945-949.

Krych-Goldberg, M. & and Atkinson, J., 2001. Structure–function relationships of complement receptor type 1. *Immunological Reviews,* 180(1), pp. 112-122.

Krych, M., Hauhart, R. & Atkinson, J. P., 1998. Structure-function analysis of the active sites of complement receptor type 1. *Journal of Biological Chemistry,* April, 273, pp. 8623-8629.

Kwan, A.H., Mobli, M., Gooley, P.R., King, G.F., Mackay, J.P., 2011. Macromolecular NMR spectroscopy for the Non-Spectroscopist. *The FEBS Journal*, January, 278, pp. 687-703.

Kwiatkowski, D. *et al.,* 1990. TNF concentration in fatal cerebral, non-fatal cerebral, and uncomplicated *Plasmodium falciparum* malaria. *Lancet,* 336, pp. 1201–1204.

Lanzillotti, R. & Coetzer, T. L., 2006. The 10 kDa domain of human erythrocyte protein 4.1 binds the *Plasmodium falciparum* EBA-181 protein. *Malaria Journal,* November, 5(100).

Laveran, C. L. A., 1893. Paludism. *The New Sydenham Society,* 46, pp. 6-8.

Lawrence, A. M. & Besir, H. U., 2009. Staining of proteins in gels with Coomassie G-250 without organic solvent and acetic acid. *Journal of Visualized Experiments,* August, 30, p. 1350.

Leslie, T. *et al.,* 2009. Epidemic of *Plasmodium falciparum* Malaria Involving Substandard Antimalarial Drugs, Pakistan, 2003. *Emerging Infectious Diseases,* November, 15(11), pp. 1753-1759.

Lin, H.H., *et al*., 2001. Molecular analysis of the epidermal growth factor-like short consensus repeat domain-mediated protein-protein interactions: dissection of the CD97-CD55 complex. *Journal of Biological Chemistry*, June, 276(26), pp. 24160-24169.

Liu, W., *et al.,* 2010. Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature,* October, 467(7314), pp. 420-425.

Lim, N. T. *et al*., 2015. Characterization of inhibitors and monoclonal antibodies that modulate the interaction between *Plasmodium falciparum* adhesin PfRh4 with its erythrocyte receptor Complement Receptor 1. *Journal of Biological Chemistry,* October, 290(42), pp. 25307-25321.

Lobo, C., Rodriguez, M., Reid, M. & Lustigman, S., 2003. Glycophorin C is the receptor for the *Plasmodium falciparum* erythrocyte binding ligand PfEBP-2 (baebl).. *Blood,* June, 101(11), pp. 4628-4631.

Lu, J., and Deutsch, C., 2001. Pegylation: a method for assessing topological accessibilities in Kv1.3. *Biochemistry*, 40(44), pp. 13288-13301.

Maier, A. G. *et al*., 2008. Exported Proteins Required for Virulence and Rigidity of *Plasmodium falciparum*-Infected Human Erythrocytes. *Cell,* July, 134(1), pp. 48-61.

Manguin, S., Bangs, M., Pothikasikom, J. & Chareonviriyaphap, T., 2010. Review on global co-transmission of human *Plasmodium* species and Wuchereria bancrofti

by Anopheles mosquitoes. *Infection, Genetics and Evolution,* March, 10(2), pp. 159-177.

Manguin, S. *et al.*, 2008. *Biodiversity of Malaria in the World.* Paris: John Libbey Eurotext.

Mata, E., *et al.,* 2013. Malaria vaccine adjuvants: latest update and challenges in preclinical and clinical research. *Biomedical Research International,* p. 282913.

Mayer, D. C. G. *et al.*, 2008. Glycophorin B is the erythrocyte receptor of *Plasmodium falciparum* erythrocyte-binding ligand, EBL-1. *Proceedings of the National Academy of Sciences of the United States of America ,* December, 106(13), pp. 5348-5352.

Meissner, M. *et al.*, 2002. A family of transmembrane microneme proteins of *Toxoplasma gondii* contain EGF-like domains and function as escorters. *Journal of Cell Science,* February, 115(Part 3), pp. 563-574.

Menéndez, C. *et al.*, 2008. An autopsy study of maternal mortality in Mozambique: the contribution of infectious diseases.. *PLoS Medicine,* February, 5(2).

Mordmüller, B. *et al.,* 2017. Sterile protection against human malaria by chemoattenuated PfSPZ vaccine. *Nature,* February, 542(7642), pp. 445-449.

Moulds, J. M. *et al.*, 2000. Identification of complement receptor one (CR1) polymorphisms in west Africa. *Genes and Immunity,* June, 1(5), pp. 325-329.

Moulds, J. M. *et al.*, 2001. Molecular identification of Knops blood group polymorphisms found in long homologous region D of complement receptor 1. *Blood,* May, 97(9), pp. 2879-2885.

Muramatsu, T., 2012. Basigin: a multifunctinoal membrane protein with an emerging role in infections by malaria parasites. *Expert Opinion on Therapeutic Targets,* October, 16(10), pp. 999-1011.

Murray, C.J.L., *et al.*, 2012. Global malaria mortality between 1980 and 2010: a systematic analysis. *The Lancet*, February, 379(9814), pp. 413-431.

Narasimham, V. & Attaran, A., 2003. Roll Back Malaria? The scarcity of international aid for malaria control. *Malaria Journal,* April, 2(8).

Newton, P.N., *et al.,* 2006. Manslaughter by Fake Artesunate in Asia—Will Africa Be Next? *PLoS Medicine*, June, 3(8), pp. 324

Ngarakana-Gwasira, E. T., Bhunu, C. P., Masocha, M. & Mashonjowa, E., 2016. Assessing the Role of Climate Change in Malaria Transmission in Africa. *Malaria Research and Treatment,* February, 2016(7104291).

Ngwa, C. J., Rosa, T. F. d. A. & Pradel, G., 2016. The Biology of Malaria Gametocytes. *Current Topics in Malaria.* November, Chapter 7.

Niang, I. *et al.*, 2014. Climate change 2014: impacts, adaptation, and vulnerability. *International Panel on Climate Change,* pp. 1199-1265.

Ntege, E. H. *et al.*, 2016. Identification of *Plasmodium falciparum* reticulocyte binding protein homologue 5-interacting protein, PfRipr, as a highly conserved blood-stage malaria vaccine candidate. *Vaccine,* November, 34(46), pp. 5612-5622.

Onyango, E. A., *et al.,* 2016. An integrated risk and vulnerability assessment framework for climate change and malaria transmission in East Africa. *Malaria Journal,* November, 15(551).

Ozbabacan, S. A., Engin, H., Gursory, A. & Keskin, O., 2011. Transient protein-protein interactions. *Protein Engineering Design and Selection,* 24(9), pp. 635-648.

Packard, R., 2008. The Making of a Tropical Disease: A Short History of Malaria. *Johns Hopkins University Press.*

Park, H. J. *et al.*, 2013. Using mutagenesis and structural biology to map the binding site for the *Plasmodium falciparum* merozoite protein PfRh4 on the human immune adherence receptor. *Journal of Biological Chemistry,* January, 289(1), pp. 450-463.

Patarroyo, M. E. *et al.*, 1987. Induction of protective immunity against experimental infection with malaria using synthetic peptides. *Nature,* August, 328(6131), pp. 629-632.

Payne, R.O, *et al.*, 2017. Human vaccination against RH5 induces neutralizing antimalarial antibodies that inhibit RH5 invasion complex interactions. *Journal of Clinical Investigations*, November, 2(21).

Periz, J., Gill, A.C., Knott, V., Handford, P.A., Tomley, F.M., 2005. Calcium binding activity of the epidermal growth factor-like domains of the apicomplexan microneme protein EtMIC4. *Molecular and Biochemical Parasitology*, October, 143(2), pp. 192-199.

Phillips, M. & Phillips-Howard, P. A., 1996. Economic implications of resistance to antimalarial drugs. *Pharmaceoeconomics,* 10(225).

Piperaki, E. & Daikos, G. L., 2016. Malaria in Europe: emerging threat or minor nuisance?. *Elsevier Clinical Microbiology and Infection,* June, 22(6), pp. 487-493.

Popper, K., 1957. Philosophy of Science: A Personal Report. *British Philosophy in Mid- Century.*

Prudêncio, M., Rodriguez, A. & and Mota, M., 2006. The silent path to thousands of merozoites: the *Plasmodium* liver stage. *Nature Reviews Microbiology,* November, 4(11), pp. 849-856.

Raja A. I *et al.* 2016. Chemically Attenuated Blood-Stage *Plasmodium yoelii* Parasites Induce Long-Lived and Strain-Transcending Protection. *Infection and Immunity,* May, 84(8), pp. 2274-2288.

Receveur-Bréchot, V., Bourhis, J., Uversky, V.N., Canard, B., Longhi, S., 2006. Assessing protein disorder and induced folding. *Proteins: Structure, Function, and Bioinformatics*, January, 62(1), pp. 24-45.

Reddy, K. *et al.*, 2015. Multiprotein complex between the GPI-anchored CyRPA with PfRH5 and PfRipr is crucial for *Plasmodium falciparum* erythrocyte invasion. *Proceedings of the National Academy of Sciences,* January, 112(4), pp. 1179–1184.

Reddy, K. S. *et al.*, 2014. Bacterially expressed full-length recombinant *Plasmodium falciparum* RH5 protein binds erythrocytes and elicits potent strain-transcending parasite-neutralizing antibodies. *Infection and Immunity,* January, 82(1), pp. 152-164.

Regev-Rudzki, N. *et al.*, 2013. Cell-cell communication between malaria-infected red blood cells via exosome-like vesicles. *Cell,* May, 153(5), pp. 1120-1133.

Renschler, J., Walters, K. M., Newton, P. N. & Laxminarayan, R., 2015. Estimated under-five deaths associated with poor-quality antimalarials in sub-Saharan Africa. *American Journal of Tropical Medicine and Hygiene ,* June, 92(6), pp. 119-126.

Richards, J. S. *et al.*, 2013. Identification and Prioritization of Merozoite Antigens as Targets of Protective Human Immunity to *Plasmodium falciparum* Malaria for Vaccine and Biomarker Development. *Journal of Immunology,* July, 191(2), pp. 795-809.

Rich, S. M. & Ayala, F. J., 2006. Malaria: Genetic and Evolutionary Aspects. *Evolutionary Origins of Human Malaria Parasites.* Supplement l, Springer US.

Rich, S. M. *et al.*, 2009. The origin of malignant malaria. *Proceedings of the National Academy of Sciences of the United States of America ,* July, 106(35), pp. 14902-14907.

Riglar, D. T. *et al.*, 2011. Super-resolution dissection of coordinated events during malaria parasite invasion of the human erythrocyte.. *Cell Host and Microbe,* January, 9(1), pp. 9-20.

Riglar, D. T. *et al.*, 2016. Localisation-based imaging of malarial antigens during erythrocyte entry reaffirms a role for AMA1 but not MTRAP in invasion.. *Journal of Cell Science,* January, 129(1), pp. 228-42.

Rodriguez, M. *et al.*, 2008. PfRH5: a novel reticulocyte-binding family homolog of *Plasmodium falciparum* that binds to the erythrocyte, and an investigation of its receptor. *PLoS ONE,* October, 3(10), p. 3300.

Rosano, G.L., and Ceccarelli, E.A., 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology,* April, 5(172).

Ross, R., 1923. Memoirs with a full account of the great malaria problem and its solution. pp. 221–239.

Rowe, A. J., Moulds, J., Newbold, C. L. & Miller, a. L. H., 1997. *P. falciparum* rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Nature,* July, 388(6639), pp. 292-295.

Rowe, J. A. *et al.*, 2000. Mapping of the region of complement receptor (CR) 1 required for *Plasmodium falciparum* rosetting and demonstration of the importance of CR1 in rosetting in field isolates. *Journal of Immunology,* December, 165(11), pp. 6341-6346.

RTS,S Clinical Trials Partnership, 2014. Efficacy and safety of the RTS,S/AS01 malaria vaccine during 18 months after vaccination: a phase 3 randomized, controlled trial in children and young infants at 11 African sites. *PLoS Medicine,* July, 11(7).

RTS,S Clinical Trials Partnership, 2015. Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial. *The Lancet,* April, 386(9988), pp. 31-45.

Russell, P. & Mohan, B. N., 1942. The immunization of fowls against mosquito-borne *Plasmodium gallinaceum* by injections of serum and of inactivated homologous sporozoites. *Journal of Experimental Medicine,* November, 76(5), pp. 477-495.

Ryan, S. J. *et al.*, 2015. Mapping Physiological Suitability Limits for Malaria in Africa Under Climate Change. *Vector-Borne and Zoonotic Diseases,* December, 15(12), pp. 718-725.

Sachs, J. & Malaney, P., 2002. The economic and social burden of malaria. *Nature,* February, 415(6872), pp. 680-685.

Sallares, R. & Gomzi, S., 2001. Biomolecular archaeology of malaria. *Ancient Biomolecules,* 3, pp. 195-213.

Sanders, P. *et al.*, 2007. Identification of protein complexes in detergent-resistant membranes of *Plasmodium falciparum* schizonts.. *Molecular and Biochemical Parasitology,* August, 154(2), pp. 148-157.

Sanders, P. *et al.*, 2005. Distinct protein classes including novel merozoite surface antigens in Raft-like membranes of *Plasmodium falciparum*. *Journal of Biological Chemistry,* December, 280(48), pp. 40169-40176.

Seder, R. A., *et al.*, 2013. Protection Against Malaria by Intravenous Immunization with a Nonreplicating Sporozoite Vaccine. *Science,* September, 341(6152), pp. 1359-1365.

Sengaloundeth, S., Green, M. D. & Fernandez , F. M., 2009. A stratified random survey of the proportion of poor quality oral artesunate sold at medicine outlets in

the Lao PDR - implications for therapeutic failure and drug resistance. *Malaria Journal,* 8(172).

Sim, B. *et al.,* 1994. Receptor and ligand domains for invasion of erythrocytes by *Plasmodium falciparum*. *Science,* June, 264(5167), pp. 1941-1944.

Singh, S.M., and Panda, A.K., 2005. Solubilization and refolding of bacterial inclusion body proteins. *J Biosci Bioeng*, April, 99(4), pp. 303-310.

Sisquella, X. *et al*., 2017. *Plasmodium falciparum* ligand binding to erythrocytes induce alterations in deformability essential for invasion. *eLife,* February, 6.

Spiegel, H. *et al*., 2015. Application of a Scalable Plant Transient Gene Expression Platform for Malaria Vaccine Development. *Frontiers in Plant Science,* December.

Spring M, M., *et al.,* 2013. First-in-human evaluation of genetically attenuated *Plasmodium falciparum* sporozoites administered by bite of *Anopheles* mosquitoes to adult volunteers. *Vaccine,* October, 31(43), pp. 4975-4983.

Spring, F. A. *et al*., 1997. The Oka blood group antigen is a marker for the M6 leukocyte activation antigen, the human homolog of OX-47 antigen, basigin and neurothelin, an immunoglobulin superfamily molecule that is widely expressed in human cells and tissues. *European Journal of Immunology,* April, 27(4), pp. 891-897.

Sreerama, N., and Woody, R.W., 2000. Estimation of protein secondary structure from CD spectra: Comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set. *Analytical Biochemistry*, 287:2, pp. 252-260.

Stack, M. L. *et al*., 2011. Estimated Economic Benefits During The 'Decade Of Vaccines' Include Treatment Savings, Gains In Labor Productivity. *Health Affairs,* June, 30(6), pp. 1021-1028.

Stanisic, D. I. & Good, M. F., 2015. Whole organism blood stage vaccines against malaria. *Vaccine,* December, 33(52), pp. 7469-7475.

Stoute, J. A. *et al.*, 1997. A Preliminary Evaluation of a Recombinant Circumsporozoite Protein Vaccine against *Plasmodium falciparum* Malaria. *The New England Journal of Medicine,* January, 336(2), pp. 86-91.

Stubbs, J. *et al.*, 2005. Molecular mechanism for switching of *P. falciparum* invasion pathways into human erythrocytes. *Science,* August, 309(5739), pp. 1384-1387.

Sturm, A. *et al.*, 2006. Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. *Science,* September, 313(5791), p. 1287–1290.

Su, X. *et al.*, 1995. The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell,* July, 82(1), pp. 89-100.

Svihrova, V. *et al.*, 2012. Costs analysis of the treatment of imported malaria. *Malaria Journal,* January, 11(1).

Tanser, F.C. *et al.*, 2003. Potential effect of climate change on malaria transmission in Africa. *The Lancet*, November, 362(9398), pp. 1792-1798.

Tamauchi, L., Coppi, A., Snounou, G. & Sinnis, P., 2007. *Plasmodium* sporozoites trickle out of the injection site. *Cellular microbiology,* 9(5), pp. 1215-1222.

Tavares, J. *et al.,* 2013. Role of host cell traversal by the malaria sporozoite during liver infection. *Journal of Experimental Medicine,* 210(5), pp. 905–915.

Tetteh-Quarcoo, P. B. *et al.*, 2012. Lack of evidence from studies of soluble protein fragments that Knops blood group polymorphisms in complement receptor-type 1 are driven by malaria. *PLoS ONE,* April, 7(4).

Tham, W.-H., Healer, J. & Cowman, A. F., 2012. Erythrocyte and reticulocyte binding-like proteins of *Plasmodium falciparum*. *Trends in Parasitology,* January, 28(1), pp. 23-30.

Tham, W. H. *et al.,* 2011. *Plasmodium falciparum* uses a key functional site in complement receptor type-1 for invasion of human erythrocytes. *Blood,* August, 118(7), pp. 1923-1933.

Tham, W. H. *et al*., 2009. Antibodies to reticulocyte binding protein-like homologue 4 inhibit invasion of *Plasmodium falciparum* into human erythrocytes. *Infection and Immunity,* June, 77(6), pp. 2427-2435.

Tham, W. *et al.,* 2015. *Plasmodium falciparum* adhesins play an essential role in signalling and activation of invasion into human erythrocytes. *PLoS Pathogens,* December, 11(12).

Tham, W. *et al.,* 2010. Complement receptor 1 is the host erythrocyte receptor for *Plasmodium falciparum* PfRh4 invasion ligand. *Proceedings of the National Academy of Sciences of the United States of America,* 107(40), pp. 17327-17332.

Tran, T.M., *et al*., 2013. Naturally Acquired Antibodies Specific for *Plasmodium falciparum* Reticulocyte-Binding Protein Homologue 5 Inhibit Parasite Growth and Predict Protection From Malaria. *The Journal of Infectious Diseases*, March, 209(5), pp. 789-798.

Triglia, T. *et al*., 2011. *Plasmodium falciparum* Merozoite Invasion Is Inhibited by Antibodies that Target the PfRh2a and b Binding Domains. *PLoS Pathogens,* June, 7(6).

Verhave, J. P., 2012. Paradigm shifts in malaria study and control. *Malaria World Journal,* 3(7).

Vinogradova, O., and Qin, J. 2012. NMR as a Unique Tool in Assessment and Complex Determination of Weak Protein–Protein Interactions. *Top Curr Chem,* June, 326, pp. 35-45

Volz, J. *et al.*, 2016. The essential invasion complex PfRh5/PfRipr/CyRPA is linked to injection of calcium into human erythrocytes by merozoites during invasion. *Cell Host Microbe,* July, 20(1), pp. 60-71.

Wanaguru, M. *et al.*, 2013. RH5–Basigin interaction plays a major role in the host tropism of *Plasmodium falciparum. Proceedings of the National Academy of Sciences of the United States of America ,* November, 110(51), pp. 20735-20740.

Wassmer, S. C. & Carlton, J. M., 2016. Glycophorins, Blood Groups, and Protection from Severe Malaria. *Trends in Parasitology,* January, 32(1), pp. 5-7.

Weaver, R. *et al.*, 2016. The association between naturally acquired IgG subclass specific antibodies to the PfRH5 invasion complex and protection from *Plasmodium falciparum* malaria. *Scientific Reports,* September, 6(33094).

Weiss, G. E., Crabb, B. S. & Gilson, P. R., 2016. Overlaying Molecular and Temporal Aspects of Malaria Parasite Invasion. *Trends in Parasitology,* April, 32(4), pp. 284-295.

Weiss, G. *et al.*, 2015. Revealing the sequence and resulting cellular morphology of receptor-ligand interactions during *Plasmodium falciparum* invasion of erythrocytes. *PLoS Pathogens,* February, 11(2).

White, N. J. *et al.*, 2009. Hyperparasitaemia and low dosing are an important source of anti-malarial drug resistance. *Malaria Journal,* November, 11(8).

Whitfield, J., 2002. Portrait of a serial killer. *Nature,* October.

Whitmore, L., and Wallace, B.A., 2004. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research*, 32, pp. W668-W673.

WHO, 2016. World Malaria Report 2016. *World Health Organisation,* p. 186.

Williams, A. R. *et al.*, 2012. Enhancing blockade of *Plasmodium falciparum* erythrocyte invasion: assessing combinations of antibodies against PfRH5 and other merozoite antigens. *PLoS Pathogens,* November, 8(11).

Wouters, M. A. *et al.*, 2005. Evolution of distinct EGF domains with specific functions. *Protein Science,* April, 14(4), pp. 1091-1103.

Wright, K. *et al.*, 2014. Structure of malaria invasion protein RH5 with erythrocyte basigin and blocking antibodies. *Nature,* November, 515(7527), pp. 427-430.

Yamana, T. K., Bomblies, A. & Eltahir, E. A., 2016. Climate change unlikely to increase malaria burden in West Africa. *Nature Climate Change,* July, 6, pp. 1009-1013.

Young, J. F., Ballou, W. R. & Hockmeyer, W. T., 1987. Developing a human malaria sporozoite vaccine. *Microbial Pathogenesis,* April, 2(4), pp. 237-240.

Young, J.F., Hockmeyer, W.T., Gross, M., Ballou, W.R., Wirtz, R.A., Trosper, J.H., Beaudoin, R.L., Hollingdale, M.R., Miller, L.H., Diggs, C.L., Rosenberg, M., 1985. Expression of *Plasmodium falciparum* circumsporozoite protein in *Escherichia coli* for potential use in a human malaria vaccine. *Science*, 228, pp. 958-962.

Young, N.S., Ioannidis, J.P.A., Al-Ubaydli, O., 2008. Why Current Publication Practices May Distort Science. *PLoS Medicine*, October, 5(10), pp. e201.

Yu, X. L. *et al.*, 2008. Crystal Structure of HAb18G/CD147 implications for immunoglobulin superfamily homophilic adhesion. *Journal of Biological Chemistry,* June, 283(26), pp. 18056-18065.

Appendix:

No.	Hybridoma	ELISA OD		No.	Hybridoma	ELISA OD
1	28/13-1A4	1.415		25	2H10	1.673
2	1A6	1.865		26	2H11	0.597
3	1A10	1.383		27	3A6	1.502
4	1E11	1.773		28	3A10	1.524
5	1H1	1.428		29	3D2	1.796
6	1H6	1.038		30	3G5	0.974
7	1H9	1.463		31	3G6	2.141
8	2A1	2.036		32	3F11	1.951
9	2A4	1.764		33	4C5	1.688
10	2A6	1.391		34	5D2	1.875
11	2A7	1.021		35	5D6	1.794
12	2A10	1.913		36	5G12	1.646
13	2A12	1.758		37	5H12	1.720
14	2B1	1.388		38	6A5	1.630
15	2E2	1.190		39	6A6	1.578
16	2E4	1.299		40	6A12	1.540
17	2E10	1.746		41	6F3	1.257
18	2E11	1.850		42	6F12	1.357
19	2H1	1.489		43	6H1	1.076
20	2H3	2.198		44	6H2	1.292
21	2H6	2.140		45	8D1	1.142
22	2H7	1.788		46	9F1	1.951
23	2H8	1.676		47	9G9	2.063
24	2H9	1.092		48	10H1	1.934
					CM	0.041
					F.S 1:6400	1.443

Table A1. ELISA Screen of PfRh4.9 Fusions: The list provided by staff at the WEHI Antibody facility of all mAbs raised against PfRh4.9 and their reactivity. Also included is the control media (CM) corresponding to sera from unvaccinated mice, and a polyclonal antibody batch of fusion sera (FS).

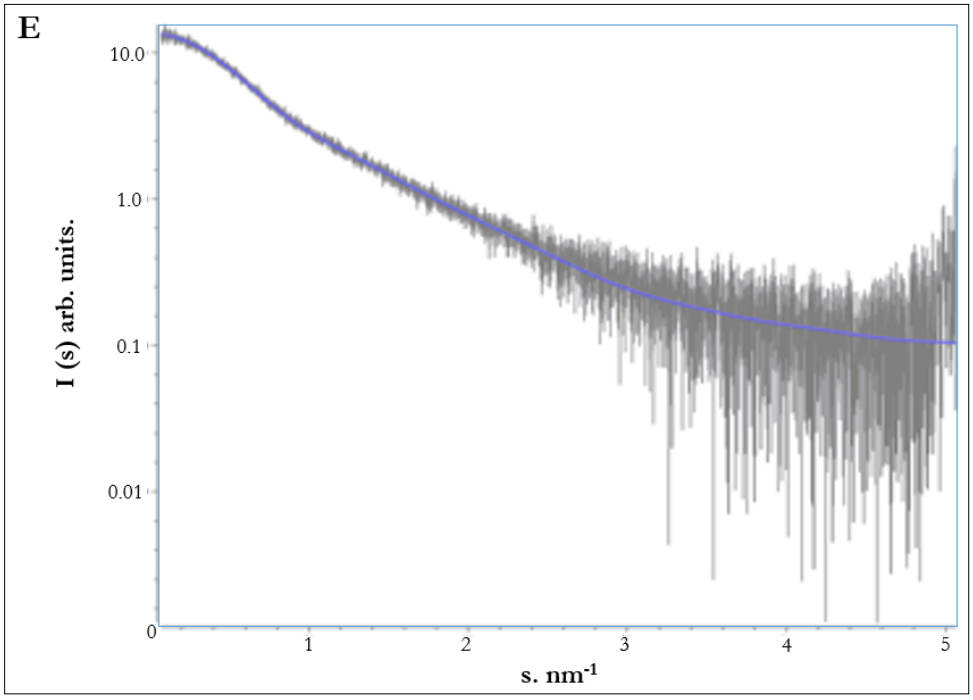


Figure A1. SAXS Data Fit for EGF5-7: The graph displays the fit to the data for the rigid body model to a χ^2 value of 1.0. CRYSOL was used to fit the best conformer of EGF5-7 (solid blue line) from homology-based modelling using SWISS-PROT and I-TASSER, and the best DAMMIF model from the SAXS data itself is given as grey lines.

Methods for the Refolding of Disulfide-Rich Proteins.

Christopher N Haggarty-Weir[1-3*], Roma Galloway[4*], and William Godfrey[5,6].

* These authors contributed equally to the work.

Author affiliations: 1. EaStCHEM, School of Chemistry, University of Edinburgh, Edinburgh, United Kingdom, 2. Walter and Eliza Hall Institute, Parkville, Victoria, 3052, Australia, 3. Department of Medical Biology, University of Melbourne, Parkville, Victoria 3052, Australia, 4. Institute for Chemical Biology, Imperial College London, South Kensington London SW7 2AZ, United Kingdom 5. Institute for Molecular Bioscience, University of Queensland, St. Lucia, Queensland 4072, Australia, 6, Institute for Genetic and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom.

Corresponding author: Christopher N Haggarty-Weir (C.N.Weir@sms.ed.ac.uk).

Author contribution: CNHW and RG carried out experiments; CNHW oversaw the teaching of the techniques and student supervision; WG contributed to the development of figures used in the manuscript; all authors contributed to the preparation of this manuscript.

Summary

- *Eschericia coli* remains a staunch workhorse for the production of recombinant proteins given its ease of handling, access and genetic manipulation using standard laboratory techniques.
- Disulfide-rich proteins can be difficult to produce in *E. coli*, in large part due to the reducing environment of the bacterial cytoplasm.
- Refolding from insoluble inclusion bodies can be a viable strategy for generating substantial quantities of disulfide-rich protein.
- In order to have the best chance of successfully refolding a protein, it is vital to carry out a variety of small-scale test refolds under a swathe of conditions.
- These conditions include altering the concentration of urea, salts, reduced and oxidized glutathione, inclusion of protease inhibitors, temperature alteration, length of refold time and protein dilution factor.
- Once a protein has undergone refolding it is vital to determine that the final product is natively folded since there is always the chance of soluble misfolded protein.
- For determination of correct folding a variety of techniques can be employed, and ideally, numerous should be used together.
- For proteins that possess enzymatic function the gold standard to assess correct folding is an activity assay. Non-enzymatic proteins can be assessed using a combination of circular dichroism and nuclear magnetic

resonance spectroscopy. These techniques should be utilized alongside mass spectrometry, Western blotting and SDS-PAGE.

#Refolding #Proteins #Biochemistry #NMR #Spectroscopy #Disulfides #Ecoli

Introduction

The first recombinant expression of a protein in *E. coli* occurred in 1976 at Genentech, where researchers produced somatostatin (Itakura K et al, 1977), thus revolutionizing the field of biotechnology and allowing for a great leap forward in biochemistry research. Since then, more advanced systems have been developed for recombinant protein production, such as the use of yeast, insect cells, plants, mammalian cells and even cell-free systems (Bill RM, 2014). But despite these advancements, *E. coli* remains a workhorse for biochemists to make proteins with. Today up to 30% of biopharmaceuticals, 50% of commercial proteins and over 70% of proteins produced in research settings are made in *E. coli* (Bill RM, 2014).

However, the system is not without its limitations. If one requires post-translational modifications, then a eukaryotic system would be preferred, and other issues such as incorrect disulfide bond formation and insoluble inclusion body production can present significant obstacles to the protein biochemist working with an *E. coli* expression system (these and other issues are reviewed in Rosano GL and Ceccarelli EA, 2014). Proteins that are rich in disulfides may incur difficulty in cysteine oxidation (which occurs in the periplasm) due to the reducing environment of the bacterial cytoplasm, which can lead to production of misfolded protein or inclusion body formation (Rosano GL and Ceccarelli EA, 2014).

Inclusion body formation may also occur due to the paucity of spatio-temporal control over foreign gene expression given that the recombinant protein is being manufactured in a potentially very different microenvironment (Rosano GL and Ceccarelli EA, 2014). There are numerous strategies to ameliorate these issues, including optimization of protein expression (i.e. altering concentration of the inducer and induction temperature), use of certain tags (such as MBP and MalE), using another organism, and of course, attempting to refold your desired protein from the inclusion bodies. Whilst challenging, it is can be possible to obtain sizeable and pure protein yields from inclusion bodies, and so carrying out a refolding trial should always be considered before abandoning *E. coli* as a production system (Burgess RR, 2009). This methods paper will outline one such refolding strategy that is scalable, suitable for disulfide-rich proteins, as well as downstream purification methods and the biophysical characterization of a final product. This method assumes use of a polyhistidine tag on the recombinant protein, however the general refolding methodology is transferable to various tags.

Methodology

Test expression

After plating freshly transformed *E. coli* containing your synthetic gene onto LB agar with a suitable selective marker (i.e. kanamycin), inoculate a colony into a 50 - 100 ml flask (preferably baffled) with 10 ml of super broth plus antibiotic, growing overnight at 37 °C with agitation. The following day prepare 9 x 300 ml baffled flasks each with 30 ml of super broth plus antibiotic and inoculate 1 ml of the starter culture into each and then place in a 37 °C incubator at 220 rpm (see fig. 1 for an overall flow-chart). Grow until an optical density at 600 nm (OD_{600}) of 0.6 – 0.8, then split the flasks so three will remain at 37 °C, three are to be placed in a 30 °C incubator and three are to be placed in an 18 °C incubator (for the latter two incubators, remove the flasks from the 37 °C incubator when the OD_{600} is just before 0.6 so as to pre-equilibrate the temperature before induction). Of each three lots of flasks, induce one of each with 0.25 mM, 0.5 mM and 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) after taking a 1 ml pre-induction sample. For the flasks at 37 °C, grow for 4 hours, take a 1 ml post-induction sample and then harvest the rest by centrifugation (8 – 10 000 rpm for 10 minutes). For the flasks at 30 °C, and 18 °C, grow for 8 and 16 hours respectively before collecting a post-induction sample, measuring the final OD_{600} and harvesting the flask contents.

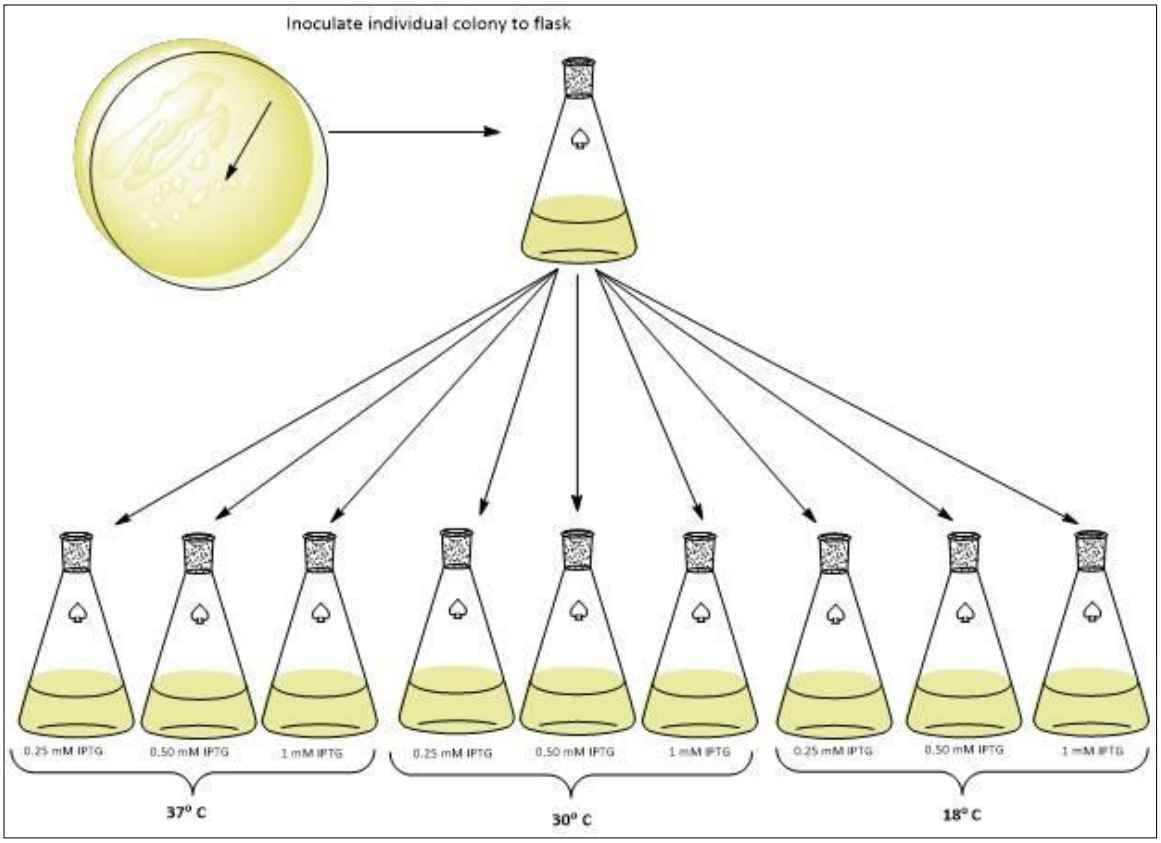


Figure 1. Flow-chart of Protein Test Expression: This figure presents the overall initial setup from post-transformed *E. coli* colony picking and starter culture, to optimization screening of both IPTG concentration and temperature.

Take 250 µl of each pre-induction sample, and 50 ul of post-induced sample, centrifuge, remove supernatant and add 10 µl of 1 x reducing sample buffer (RSB), then boil for 1 minute at ~97 °C. Carry out an SDS-PAGE analysis on these samples. For each sample showing evidence of induction for recombinant protein production, centrifuge and resuspend in 2 ml lysis buffer (50 mM Tris pH 8, 250 mM NaCl, 0.1% triton X-100, 2 mM EDTA, spatula tip of lysozyme and 1.5 ul of DNaseI), then process via sonication (90 sec on/100 sec off at 35% amplitude for 7.5 minutes) or mechanical cell disruption. Take a 50 ul sample post-lysis and centrifuge at 13 000 rpm for 1 minute, then separate the supernatant from the pellet. Take 6 ul of the supernatant and add 10 ul of 1 x RSB; for the pellet, add 50 ul of water, resuspend thoroughly so that 6 ul can be taken and added to 10 ul of 1 x RSB. Boil samples for 1 minute at ~97 °C. Carry out an SDS-PAGE analysis on these samples to determine that you have protein expressing in the inclusion body fraction.

Protein production scale-up and inclusion body production

Set up 1 x 500 ml of your culture media broth with antibiotic (per construct) and inoculate each with a starter culture grown overnight from your transformed

bacteria. Grow at the temperature and induce with an IPTG concentration as previously determined. Process as described in the “test expression” section (adding protease inhibitors such as PMSF or a protease inhibitor cocktail at the cell lysis step, and scaling up the volumes).

Once you have a pellet post-lysis, carry out 2 - 3 more washes with 25 mls of lysis buffer and if possible, sonicate between each wash before spinning the pellet down. Weigh out the pellet and if you need to store, do so at -20 °C. Solubilize the pellets with 5 ml/g of solubilization buffer (5 M guanidine HCl, 250 mM NaCl, and 20 mM Tris pH 8, and 20 mM Beta-mercaptoethanol βME). Leave this on a rotor for over an hour at room temperature then centrifuge at 9000 rpm for 10 mins. Transfer supernatant to a tube with 3 mls settled Ni-NTA resin (assuming protein has a poly-His tag). Leave this on the rotor overnight at room temperature.

Affinity purification

Let the resin settle and set up disposable tubes for the resin and sample. Add most of the fluid above the resin to the columns and collect the first 2 CVs (6 ml) as breakthrough fractions (BT) and collect the rest in a separate tube. Now mix the resin and add to the column, allowing it to run empty into the aforementioned tube. Get fresh collection tubes, add 10 CVs (30 ml) of solubilization buffer and collect as the guanidine wash fraction. Add 10 CV of 8M Urea wash/concentration buffer (8 M urea, 250 mM NaCl, 20 mM Tris pH 8 and 20 mM βME) and collect. Add 10 CVs of elute buffer with 20 mM BME and use 20 x 1.5 ml tubes to collect fractions.

Add 5 x 0.5 CV of elute buffer (8 M urea, 250 mM NaCl, 20 mM Tris pH 8, 1 M imidazole 20 mM βME), mix and leave for 15 minutes, elute and collect in tubes. Add 10 CV of solubilization buffer and wash through (collect in a tube), then add 5 ml more, mix and reapply the breakthrough (after taking samples for gels) and re-bind overnight. You may get numerous batches of protein this way.

Run samples on a gel via SDS-PAGE; either take neat samples if you think there is significant protein (measure Abs_{280}), or carry out a TCA precipitation. Combine all fractions that have protein in them and aliquote out for the refold test.

Refolding trials.

Take 1 ml of your eluted protein sample and add 100 ul of 200 mM DTT (made in solubilization buffer without βME) and call this your reducing sample (R). Take 1 ml of another aliquote and add 100 ul solubilization buffer to make your Non-Reducing sample (NR). Leave at room temperature for 2 hours.

In a tube, prepare 4 mls of 0.1 M reduced glutathione, and in another make up 500 ul of 0.1 M oxidized glutathione. Get 8 x 50 ml tubes and label half R and half NR. Of the R half, label each one as follows (then do the same for the NR tubes)- 0 M Urea, 1 M Urea, 2 M Urea, 3 M Urea. These tubes are to hold 20 mls of solution by the end of the set up.

Add urea salts to the tubes (0 g for the 0 M tubes, 1.2012 g for the 1 M tubes, 2.4024 g for the 2 M tubes and 3.6036 g for the 3 M tubes). Next, add 200 ul of the reduced glutathione solution to each, then 20 mM Tris pH 8 (pH depends on the isoelectric point of your protein; the proteins used in this paper were all around was 5.25 - 5.75) and 100 mM NaCl. Top the tubes up to 20 mls with purified water and adjust pH. See fig. 2 for an overall schematic of the steps up to this point.

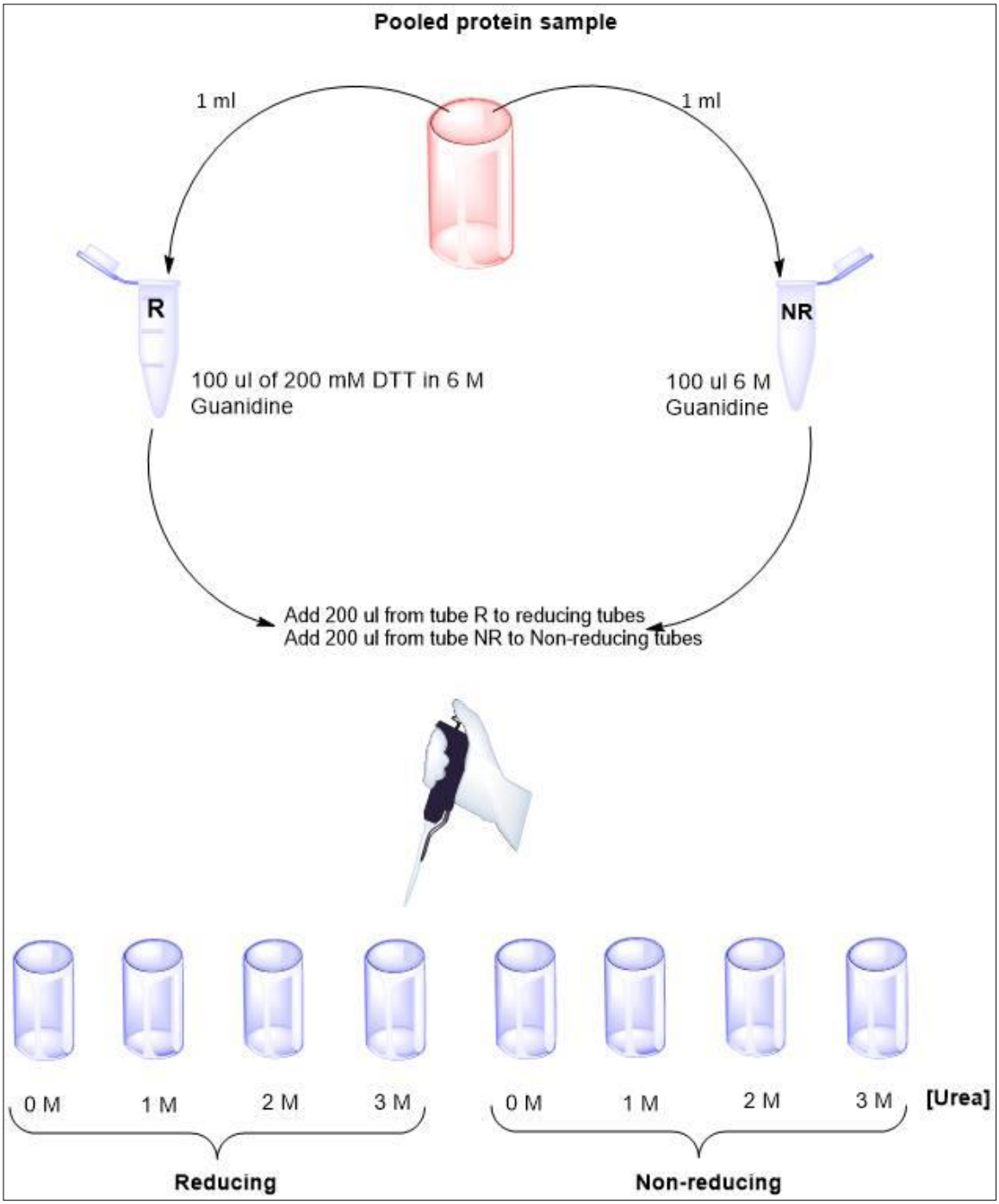


Figure 2. Schematic of Protein Refolding Set-up: The pool of Ni-affinity purificed fractions of protein should be used for the optimization set-up of the various refolding conditions (R or NR and concentration of Urea).

Slowly add 200 ul of your protein to each tube in a slow drop-by-drop fashion. Add 200 ul of the oxidized glutathione solution to each tube. Take a 2 ml sample

from each tube and carry out a TCA precipitation; this will be your t_0 sample and you should have enough to run an R and NR sample next to each other on a gel.

Leave the tubes mixing at room temperature. Each day, for 3 days, take a 2 ml sample from each tube, carry out a TCA and analyse R and NR samples. By the end of these trials you will see if the addition of urea is necessary. Further you will look for the conditions that give you the best collapse of bands over time. Note; you may not need to do a TCA if you have a great deal of protein. It is worth taking 30 ul samples and running them neat on reducing gels. Once you have the above conditions worked out, then you need to optimize the levels of glutathione used. Try the following- 1 mM R glut/1 mM Ox glut, 1 mM R glut/ 0.5 mM Ox glut, and 1 mM R glut/ 0.25 mM Ox glut.

By now you should know the conditions for scaling up the refold. The final major step that can be optimized here is the dilution factor (i.e. how much protein goes into how much refold buffer). It is good to start off with 1:100, but also try 1:50 and 1:25. Also, generally it is good not to go too large with the bottle size. Only use 250 ml bottles with 250 ml of buffer (don't use 1 L bottles with 1 L buffer unless you want to attempt to optimize this).

Scale up.

(Note, this assumes a 1:25 dilution factor so if you are using 1:100 just alter the amount of protein being added to the mix).

Make up your 20 mM Tris/100 mM Nacl buffer, filter then chill on ice whilst bubbling through nitrogen gas for 20 mins. Get clean 250 ml bottles (the amount depends on how much protein you plan to make, often it is good only to start with 2 bottles until you have successfully made 1 batch) each with clean magnetic stirrers in them. Fill with 250 ml of the aforementioned Tris/Nacl buffer, stir in the cold room and add a protease inhibitor tablet into each.

Add 100^{th} of the volume (2.5 mls) of the stock 0.1 M reduced glutathione (if this is the concentration you determined was best) to each bottle. Next, add 10 mls of your protein to each bottle (if this is your first time you might want to try 2.5 mls) in a very slow, drop-wise fashion, whilst the magnetic stirrer is on. Finally, add 2.5 mls of stock oxidized glutathione (concentration will depend on what you previously worked out during optimization). Aerate the top of each bottle with nitrogen gas for 1 min, put cap on and parafilm. Leave stirring at room temp for 3 days (or the amount of days you optimized before; additionally you can try carrying the refold out at 4 °C, or without the magnetic stirrer). You can take R and NR samples each day to track the refold.

Dialysis.

After taking pre-dialysis samples to run R and NR SDS-PAGE gels of, put 2 x 250 ml refold into dialysis tubes of a suitable molecular weight cut-off and then these into a 5000 ml beaker with 20 mM Tris pH 8 and 20 mM NaCl, with two buffer changes a day for 2 days. Carry out the dialysis at 4 °C and place a magnetic stirrer in the bottom of the large beaker.

Take samples out of dialysis bags, centrifuge at top speed to remove particulates (filtering can work, but beware if your protein non-specifically binds the filter). Now take your post-dialysis samples for SDS-PAGE.

Ion exchange and size exclusion chromatography

The final step in the purification is to obtain a purified protein sample for biophysical characterization to assess the folding of your protein (i.e. nuclear magnetic resonance, circular dichroism, mass spectrometry, and/or Western blotting) or an enzymatic activity assay if suitable for your protein. Disulfide mapping can also be extremely informative and an example of a methodology is provided by Hodder AN et al., 1996. This is beyond the scope of this methods paper, however in general, one can first utilize ion exchange followed by size exclusion chromatography to obtain highly pure samples. An example of how the sample should look at key stages in the purification process if shown in fig. 3.

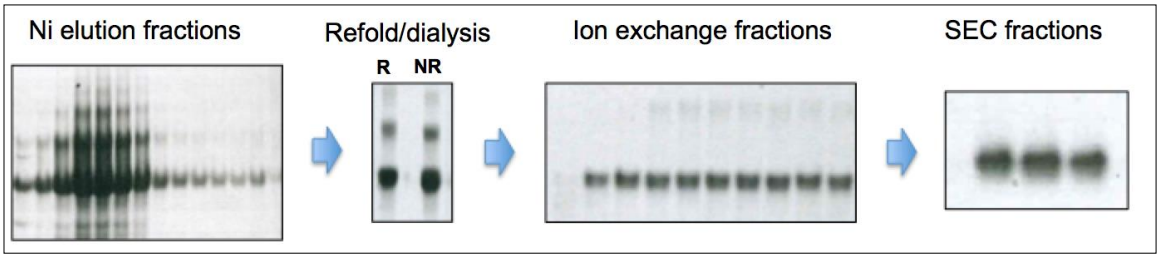


Figure 3. Protein purification steps: The Ni-affinity fractions eluted from the column can be run under R conditions on SDS-PAGE where both an analysis of the size of the protein can be assessed (shown as the predominant band), in addition to the higher-order species. During refolding the disulfides should shuffle and reorganize into lower order species, but generally will still include monomer, dimer and trimer. Ion exchange chromatography can be used to obtain monomer and dimer, however a final purification step may be needed to get pure monomeric protein (here demonstrated by post-SEC samples).

Final Comments

This method, whilst not all-inclusive of potential refolding methods, should provide the biochemist with a reasonable level of options for trialing the refolding of their inclusion bodies. Many other conditions can be attempted, including the

temperature that the refold is occurring at, or the use of L-arginine buffers. Of course, if one experiences little progress in attempting refolding, they could contact the authors for further options or move to another expression system, such as baculovirus or *Pichia*.

Acknowledgements

All authors would like to thank the people who contributed to the Pozible malaria project run by the Walter and Eliza Hall Institute, which contributed to the work shown here. CNHW would like to thank Dr. Anthony N Hodder for the training he received in biochemistry, and the staff of Andrew Usher & Co. in Edinburgh for their service whilst this manuscript was being prepared.

Conflicts of Interest

The authors state no conflicts of interest.

Grant Information

Funding from the Pozible malaria funds (project #190553), the Australian Government via the Australian Postgraduate Award scheme and the Australia - Europe Malaria Research Cooperation & the Australian Society for Parasitology travel award funds all contributed to the work mentioned in this paper.

References

Bill RM. Playing catch-up with Escherichia coli: using yeast to increase success rates in recombinant protein production experiments. Frontiers in Microbiology. 2014 Mar 5;(85).
http://journal.frontiersin.org/article/10.3389/fmicb.2014.00085/full

Burgess RR. Refolding solubilized inclusion body proteins. Methods in Enzymology 2009, 463;259-82.
http://www.ncbi.nlm.nih.gov/pubmed/19892177

Hodder AN, et al. The disulfide bond structure of *Plasmodium* apical membrane antigen-1. Journal of Biological Chemistry 1996 Nov 15;271(46).
http://www.jbc.org/content/271/46/29446.abstract

Itakura K, Hirose T, Crea R, Riggs AD, Heynecker HL, Bolivar F et al. Expression in eschericia coli of a chemically synthesized gene for the hormone somatostatin. Science. 1977 Dec 9;198(4321):1056-63.

http://science.sciencemag.org/content/198/4321/1056.long

Rosano GL and Ceccarelli EA. Recombinant protein expression in Escherichia coli: advances and challenges. Frontiers in Microbiology 2014 Apr 17; 5;(172). http://journal.frontiersin.org/article/10.3389/fmicb.2014.00172/full

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 290, NO. 42, pp. 25307–25321, October 16, 2015

Characterization of Inhibitors and Monoclonal Antibodies That Modulate the Interaction between *Plasmodium falciparum* Adhesin PfRh4 with Its Erythrocyte Receptor Complement Receptor 1*

Received for publication, April 8, 2015, and in revised form, July 24, 2015 Published, JBC Papers in Press, August 31, 2015, DOI 10.1074/jbc.M115.657171

Nicholas T. Y. Lim[‡], Markus J. Harder[§], Alexander T. Kennedy[‡¶], Clara S. Lin[‡¶], Christopher Weir[‡¶‖], Alan F. Cowman[‡¶], Melissa J. Call[‡¶], Christoph Q. Schmidt[§], and Wai-Hong Tham[‡¶1]

From the [‡]Walter and Eliza Hall Institute, Parkville, Victoria 3052, Australia, the [§]Institute of Pharmacology of Natural Products and Clinical Pharmacology, Ulm University, Helmholtzstrasse 20, D-89081 Ulm, Germany, the [¶]Department of Medical Biology, University of Melbourne, Parkville, Victoria 3052, Australia, and the [‖]School of Chemistry, University of Edinburgh, Edinburgh EH93JJ, Scotland, United Kingdom

Background: PfRh4 binds complement receptor 1 to mediate malaria parasite entry into red blood cells.
Results: Monoclonal antibodies and inhibitors either block or enhance PfRh4 interaction with complement receptor 1.
Conclusion: Identification was made of critical regions and residues within PfRh4 and CR1 that mediate successful *P. falciparum* entry.
Significance: Understanding functional regions within PfRh4 will aid in design of vaccine subunits.

***Plasmodium falciparum* parasites must invade red blood cells to survive within humans. Entry into red blood cells is governed by interactions between parasite adhesins and red blood cell receptors. Previously we identified that *P. falciparum* reticulocyte binding protein-like homologue 4 (PfRh4) binds to complement receptor 1 (CR1) to mediate entry of malaria parasites into human red blood cells. In this report we characterize a collection of anti-PfRh4 monoclonal antibodies and CR1 protein fragments that modulate the interaction between PfRh4 and CR1. We identify an anti-PfRh4 monoclonal that blocks PfRh4-CR1 interaction *in vitro*, inhibits PfRh4 binding to red blood cells, and as a result abolishes the PfRh4-CR1 invasion pathway in *P. falciparum*. Epitope mapping of anti-PfRh4 monoclonal antibodies identified distinct functional regions within PfRh4 involved in modulating its interaction with CR1. Furthermore, we designed a set of protein fragments based on extensive mutagenesis analyses of the PfRh4 binding site on CR1 and determined their interaction affinities using surface plasmon resonance. These CR1 protein fragments bind tightly to PfRh4 and also function as soluble inhibitors to block PfRh4 binding to red blood cells and to inhibit the PfRh4-CR1 invasion pathway. Our findings can aid future efforts in designing specific single epitope antibodies to block *P. falciparum* invasion via complement receptor 1.**

Malaria parasites are obligate intracellular microbes exquisitely adapted for invasion and survival within the red blood cell of the host. To gain entry into red blood cells, the parasite must recognize and bind to these cells as well as activate a complex series of steps that involve multiple protein-protein interactions between parasite and host cells (for review see Ref. 1). The subsequent cycles of growth, replication of parasites, and egress from infected blood cells are responsible for the symptoms associated with malaria (for review see Ref. 2).

Parasite invasion begins with initial recognition and attachment of merozoites, the invasive form of malaria parasites, to red blood cells (3). This interaction is dynamic and involves considerable deformation of the red blood cell membrane as the parasite rolls across the host cell surface (4). After initial attachment, the merozoite orientates itself to juxtapose its apical prominence with the red blood cell surface. This allows parasite adhesins localized at the apical tip to interact with their cognate receptors to mediate irreversible attachment and commitment to invasion (5, 6). After this, a tight junction is formed between the parasite and the red blood cell membrane (7). Active invasion proceeds through an invagination of the surface and the tight junction moves from the apical to posterior pole of the merozoite, powered by the parasite's acto-myosin motor. Once the merozoite is inside the red blood cell, the red blood cell membrane is sealed behind it, completing invasion. Remarkably, the entire invasion process is accomplished within a few minutes (4).

Of the five human malaria species, *Plasmodium falciparum* is the most lethal. In *P. falciparum*, two gene families encode important parasite adhesins utilized for engagement with red blood cell receptors: erythrocyte binding-like antigens (EBAs; EBA175, EBA181, EBA140, EBL-1) (8, 9) and *P. falciparum* reticulocyte binding-like homolog proteins (PfRhs; Rh1, Rh2a/b, Rh4, Rh5) (10–12). During invasion these adhesins localize to the apical tip of the merozoite and bind specific receptors to initiate parasite entry into human red blood cells. Several red blood cell receptors have been identified as entry

* This work was supported by National Health and Medical Research Council Grant APP1026581 and an Australian Research Council Future Fellowship (to W. H.-T.). The authors declare that they have no conflicts of interest with the contents of this article.
[1] To whom correspondence should be addressed: Division of Infection and Immunity, Walter and Eliza Hall Institute, 1G Royal Parade, Parkville, Victoria 3052, Australia. Tel.: 61-3-93452716; Fax: 61-3-93470852; E-mail: tham@wehi.edu.au.

points for *P. falciparum*. The glycophorins were the first identified and exclusively bind members of the PfEBA[2] family of proteins (glycophorin A to EBA-175, glycophorin B to EBL-1, and glycophorin C to EBA-140) (13–16). For the PfRh family of proteins, only two cognate adhesion-receptor pairs have been identified: PfRh4 to complement receptor 1 (CR1) and PfRh5 to basigin (17–19).

Molecular studies show that EBA and PfRh proteins are potential vaccine candidates because: (i) they play a crucial role in merozoite invasion; (ii) anti-EBA and anti-PfRhs antibodies inhibit parasite invasion; (iii) epidemiological studies show that most anti-EBA and anti-PfRhs antibodies in human cohort studies correlate with protection from clinical disease (1, 20–25). In addition to clinical applications, monoclonal antibodies (mAbs) have provided unique tools with which to study molecular pathways of parasite invasion, and some of the functional regions of PfRh family members have been recently characterized. Anti-PfRh1 mAbs affect rhoptry secretion and calcium signaling after merozoite attachment to erythrocytes, suggesting that PfRh1 plays a role in downstream signaling events (26). Anti-PfRh5 antibodies that interfere with the PfRh5-basigin interaction completely inhibit the ability of *P. falciparum* parasites to invade red blood cells (23, 27), and the mechanism by which inhibition occurs has been elucidated by the recent crystal structures of PfRh5 alone and with either its receptor basigin or neutralizing antibodies (28, 29). PfRh5 adopts a novel fold using a α-helical scaffold that provides binding sites at the tips of helices for basigin and some inhibitory monoclonal antibodies (29). The high resolution structures of PfRh5-basigin and PfRh5-mAbs binding interfaces will clearly allow future structure-guided design of inhibitory epitopes for more potent neutralizing mAbs.

Characterization of the PfRh4-CR1 invasion pathway has validated the potential of PfRh4 as a vaccine candidate (for review see Ref. 30). A soluble fragment of the PfRh4 ectodomain (rPfRh4) that encompasses the red blood cell binding region can be successfully expressed in *Escherichia coli*, and rabbit polyclonal antibodies raised against this fragment are able to inhibit PfRh4 binding to red blood cells (31). Furthermore, affinity-purified anti-rPfRh4 human antibodies from individuals in malaria endemic regions inhibited *P. falciparum* invasion via the PfRh4-CR1 pathway and correlated with protection (24). Immunization (in rabbits) with a combination of EBA-175, PfRh2a/b, and PfRh4 recombinant proteins induced antibodies that potently blocked merozoite invasion *in vitro* (22).

Previous work has mapped the PfRh4-interacting region on CR1 and also identified soluble forms of CR1 that are able to act as competitive inhibitors (Refs. 17, 31, and 32 and reviewed in Ref. 30). CR1 is a type one integral membrane glycoprotein composed of an N-terminal ectodomain that has a number of allelic variants, a transmembrane region, and a C-terminal cytoplasmic domain. The most common allelic variant of CR1 is composed of 28–30 structural modules called complement control protein (CCP) modules in the extracellular domain. A truncated form of CR1 (sCR1) lacking the transmembrane and cytoplasmic domain, inhibits PfRh4 binding to CR1 on the red blood cell surface (17). Clinical isolates from Kenya also demonstrated a significant utilization of CR1 for invasion of intact erythrocytes that was inhibited in the presence of sCR1 (34). Initial mapping studies identified the first three modules of CR1 (CCPs 1–3) as the most specific inhibitor of the PfRh4-CR1 invasion pathway (33). Recent work using CCPs 1–3 helped define the role of PfRh4 in the deformation of red blood cell membrane during *P. falciparum* invasion into red blood cells (35). Further mapping of the PfRh4 binding site on CR1 using truncation and deletion constructs pinpoint CCP 1 as the major binding site for PfRh4, and extensive mutagenesis experiments within this domain clearly delineated the PfRh4 binding site (32). These studies employed ELISA, co-immunoprecipitation and surface plasmon resonance (SPR) to characterize mutations that affected PfRh4-CR1 complex formation and showed that clustered mutations in residues 6–9 or single mutations in residues 18 and 20 resulted in a dramatic loss in affinity for rPfRh4. Park *et al.* (32) were able to engineer an artificial binding site within CCPs 8–14 by substituting residues within CCP 1 that are critical for PfRh4 interaction to their homologous position in CCP 8. Strikingly, this engineered site within CCPs 8–14 showed a 30-fold higher affinity for rPfRh4. Although the effects of the mutations are well understood in biochemical protein-protein interaction assays, it will be important to determine in a cellular context if any of these mutations lose their ability to block PfRh4-CR1 invasion or, in the case of the engineered site, lead to a potentially better inhibitor of *P. falciparum* invasion.

The availability of anti-PfRh4 mAbs that interfere with the PfRh4-CR1 interaction would provide an important tool in the identification of inhibitory epitopes in the binding interface. In this paper we generated anti-PfRh4 mAbs and tested their ability to modulate the interaction between PfRh4 and CR1 *in vitro* and to inhibit *P. falciparum* invasion. Furthermore, we characterize a collection of CR1-based inhibitors that will be invaluable in determining structure-function relationships between this ligand-receptor pair. Our results will identify distinct functional regions within PfRh4 and CR1 that are important for mediating entry of *P. falciparum* parasites into human red blood cells.

Experimental Procedures

Anti-PfRh4 Mouse Monoclonal Antibodies Production—Anti-PfRh4 mAbs were produced at the Monoclonal Antibody Facility at the Walter and Eliza Hall Institute. BALB/c and C57Bl6 mice received three immunizations of recombinant PfRh4 purified as described below. At day 0, Complete Freund's adjuvant was mixed with the antigen into an emulsion and injected intraperitoneally. At day 30 and day 60 the antigen was mixed with incomplete Freund's adjuvant, and the emulsion was injected intraperitoneally. Serum ELISA titrations were performed at day 72. The mouse with the best response received a final injection of antigen in saline, and splenocytes

[2] The abbreviations used are: EBA, erythrocyte binding-like antigen; PfRh, *P. falciparum* reticulocyte binding-like homolog protein; rPfRh, recombinant PfRh; CR1, complement receptor 1; mAb, monoclonal antibody; CCP, complement control protein; SPR, surface plasmon resonance; Bis-Tris. 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol.

were harvested 3 days later. Spleen cells were fused with SP2/0 myeloma cells to form B-cell-myeloma fused cells (hybridomas). Hybridomas were grown in hypoxanthine-aminopterin thymidine growth medium. ELISA was used to select hybridomas producing antibodies specific to PfRh4. Hybridomas were cloned by limiting dilution in multiwell plates aiming for one cell or less per well. The subcloning supernatants were screened by ELISA. Two or more rounds of limiting dilution cloning were generally required before the hybridomas were deemed monoclonal. The antibodies were purified from monoclonal hybridoma supernatants with protein A-Sepharose.

ELISA—96-Well flat-bottomed plates (Maxisorp; Nunc) were coated with rPfRh4 (1 μg/well) resuspended in PBS and incubated for 2 h at room temperature. Plates were incubated with 5% skim milk, 0.01% Tween 20 for 1 h at room temperature to block nonspecific binding. After washing, anti-PfRh4 mAbs were added at 1:1000 dilution for 1 h at room temperature. Plates were washed 3 times before the addition of HRP-conjugated goat anti-mouse secondary antibodies (1:1000 dilution) for 1 h at room temperature. Azino-bis-3-ethylbenthiazoline-6-sulfonic acid (ABTS liquid substrate; Sigma) was used to detect HRP activity. 1% SDS was used to stop the reaction, and absorbance was measured at 405 nm. All washes were done in PBS, 0.01% Tween 20, and dilutions of antibodies were done in 0.5% skim milk, 0.01% Tween 20. All samples were performed in duplicate.

For competitive binding experiments with mAbs, the ELISA protocol was performed as above with the following modifications. Microtiter wells were coated with CCPs 1–3 (1 μg/well). After washing and blocking, rPfRh4 was added at a final concentration of 1 μg/well together with anti-PfRh4 mAbs. Binding of rPfRh4 was detected using an anti-PfRh4 rabbit polyclonal antibody followed by addition of HRP conjugated goat anti-rabbit secondary antibody.

Indirect Immunofluorescence Assay—Indirect immunofluorescence assay samples were prepared from magnet-purified parasites as described previously (36). Briefly, parasites were smeared on glass slides and fixed using 100% ice-cold methanol for 30 s. After fixation, samples were blocked overnight in 3% BSA (Sigma) in PBS at 4 °C. Primary antibodies were diluted in blocking solution and applied to slides. After a 1-h incubation at room temperature, the slides were washed 3 times in PBS. This was followed with a 1-h incubation with secondary antibodies diluted in blocking solution; slides were protected from light during the incubation. The slides were washed three times with PBS and mounted in VectaShield (Vector Laboratories) with 0.1 ng/μl 4′,6-diamidino-2-phenylindole (DAPI, Invitrogen). Primary antibodies used in indirect immunofluorescence assay imaging were diluted in 3% BSA in PBS, and the concentrations were as follows: mouse anti-PfRh4 (2 mg/ml, 1:500), rabbit anti-PfRh2 (2 mg/ml, 1:500). Secondary antibodies were used at the following concentrations: Alexa Fluor 488 goat anti-mouse (1:500) and Alexa Fluor 594 goat anti-rabbit (1:500). Fluorescence images were obtained using DeltaVision Elite widefield fluorescence microscope. Z stacks were taken above and below parasites and processed using Axiovision deconvolution software package.

Immunoblotting and Antibodies—Proteins larger than 100 kDa were run on 3–8% Tris acetate, whereas smaller proteins were run on 4–12% Bis-Tris SDS-PAGE gels (Invitrogen). Western blotting was performed using standard protocols, and the blots were processed with an enhanced chemiluminescence (ECL) system (Amersham Biosciences). Western blots probed with anti-PfRh4 mouse mAbs were conducted 2 mg/ml at 1:1000 dilution followed by HRP-conjugated goat anti-mouse secondary antibody at a 1:1000 dilution.

Immunoprecipitation Assays—In a reaction volume of 100 μl, anti-PfRh4 mAbs were incubated with recombinant PfRh4 at 0.1 and 0.05 mg/ml, respectively, for 1 h at room temperature. 5 μl of packed Protein G-Sepharose beads were added to capture the anti-PfRh4 mAbs; this incubation proceeded for 1 h at room temperature. Beads were washed five times with PBS, and proteins were eluted with equivalent volumes of 2× nonreducing sample buffer and boiled for 5 min before separating on SDS/PAGE gels. Protein eluates were visualized using SimplyBlue SafeStain (Life Technologies), performed according to the manufacturer's protocol.

Fluorescence Resonance Energy Transfer (FRET) Assay—To screen for antibodies that disrupt the PfRh4 interaction with CR1, we labeled rPfRh4 and the CCPs 1 and 2 with NHS ester derivatives of DyLight 594 and DyLight 488 (Life Technologies), respectively, to assess the rPfRh4-CCPs 1 and 2 interaction by FRET. rPfRh4 (2 mg/ml) was labeled at a 1:3 molar ratio with DyLight 594 in 50 mM MES, pH 6.0, 150 mM sodium chloride. Buffer exchange to 50 mM MES pH 6.0 and free dye removal was achieved using a Micro Bio-Spin P-6 column (Bio-Rad). CCPs 1 and 2 were similarly labeled with DyLight 488, and each preparation was assessed by spectrophotometry to determine the average number of dye molecules per protein molecule. After labeling, each rPfRh4 molecule carried an average of 2.6 DyLight-594 molecules, whereas CCPs 1 and 2 carried 1.8 DyLight 488 molecules. 10-μl binding reactions of rPfRh4·DyLight-594 and CCPs 1 and 2·DyLight488, each at 125 nM in 50 mM MES, pH 6.0, 150 mM sodium chloride, were placed in Corning 384-well plates (#3820), and fluorescence intensity at the following wavelengths were measured using an EnVision plate reader (PerkinElmer Life Sciences): DyLight-488 (donor) 485/14-nm excitation filter and 535/25-nm emission filter, DyLight-594 (acceptor) 590/20-nm excitation filter and 615/9-nm emission filter, and sensitized emission 485/14-nm excitation filter and 615/9-nm emission filter. In an ideal situation sensitized emission should only be measured when DyLight-488 and DyLight-594 are in close proximity, but in practice we measured a significant contribution to FRET-mediated sensitized emission from DyLight-594 in the absence of DyLight-488 and chose to correct for this by calculating FRET as a ratio of sensitized emission over DyLight-594 fluorescence intensity. This creates a unit-less FRET ratio that can be used to compare binding of rPfRh4·DyLight-594 to CCPs 1 and 2·DyLight-488 in the presence or absence of candidate inhibitors that disrupt the interaction.

Flow Cytometry-based Erythrocyte Binding Assay—The flow cytometry-based erythrocyte binding assay was performed as described with the following modifications (17). 40 μl of packed erythrocytes were washed twice with 400 μl of 1% BSA/PBS and

resuspended to a final volume of 1×10^7 cells/ml. Meanwhile, 1.25 μg of rPfRh4 was preincubated with 10 μg of mAb at room temperature for 10 min. 100 μl of the resuspended erythrocytes was added to the recombinant protein/antibody mix and incubated for 30 min. After binding, erythrocytes were washed three times with 1% BSA, PBS. To detect rPfRh4 binding, anti-PfRh4 rabbit IgG R550 (1:100 dilution) was added and incubated for 30 min followed by 3 washes with 1% BSA, PBS. Alexa Fluor 488 goat anti-rabbit secondary antibody (1:100, Life Technologies) was added and incubated at room temperature and protected from light for 30 min. Erythrocytes were washed twice with 1% BSA. PBS followed by 1 wash in PBS and resuspended in 500 μl of PBS before being transferred to FACS tubes. A total of 50,000 erythrocytes were read on the FACSCalibur flow cytometer (BD Biosciences), and the results were analyzed using FlowJo software (Three Star). The percentage of erythrocytes with bound rPfRh4 was determined by normalizing the number of erythrocytes exhibiting a positive Alexa Fluor 488 signal that is above the background (which is Alexa Fluor 488 signal of erythrocytes without rPfRh4 added) on the total number of erythrocytes.

Parasite Culture and Growth Inhibition Assay—*P. falciparum* asexual stages were maintained in human O+ erythrocytes in RPMI-HEPES medium with 50 μg/ml hypoxanthine, 25 mM $NaHCO_3$, 20 μg/ml gentamicin, and 0.5% Albumax II (Gibco; Invitrogen) in 1% O_2, 4% CO, and 95% N_2 at 37 °C and synchronized by standard methods. We used the D10-PHG for the majority of our assays, which are GFP-positive (37).

Growth inhibition assays were performed as described with the following modifications (33). Starting parasitemia for the growth assays was 0.2% with a hematocrit of ~1%. CR1 constructs, anti-PfRh4 mAbs, bovine serum albumin (BSA, Sigma), and heparin were added at the beginning of the assay. Growth inhibition assays were performed in 96-well round-bottom microtiter plates (BD Biosciences) over 2 cycles of parasite growth. After 96 h the parasitemia was determined by flow cytometry of GFP-positive and EtBr-stained trophozoite-stage parasites using a FACSCalibur (BD Biosciences) and a plate reader. For each well 40,000 cells or more were counted. Growth was expressed as mean parasitemia obtained from triplicate readings. At least two independent assays were performed, each in triplicate. % Growth refers to the % parasitemia in the presence of CR1 constructs or anti-PfRh4 mAbs relative to the % parasitemia with the addition of PBS (no mAb control, which was arbitrarily set to be 100%).

Mapping of mAb Epitopes—An overlapping peptide array consisting of 15-mer biotinylated peptides with a SGSG linker was synthesized by Mimotopes Pty Ltd, Clayton, Australia. Peptides were resuspended and applied to streptavidin-coated plates. ELISA was performed as stated above.

Antibody Affinity Measurements—Binding of rPfRh4 to anti-PfRh4 mAbs 5H12 and 10C9 was monitored by SPR using the Biacore 3000 system (GE Healthcare), where experiments were performed at 25 °C in PBS-*p* +. CMD-500 (XanTec bioanalytics GmbH, Dusseldorf, Germany) chip surfaces were prepared by immobilization of anti-mouse IgG at 30 μg/ml in 10 mM sodium acetate, pH 5.0, using standard NHS-EDC (EDC, 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide) amine coupling methods, where a final immobilization of 7542.2 resonance units was achieved. The antibody affinity assay was performed using an indirect capture method where antibodies 5H12 and 10C9 were diluted to 30 μg/ml in PBS-*p* + and injected for 120 s at 5 μl/min with a 300-s stabilization time. After stabilization, rPfRh4 in PBS-*p* + was injected at 20 μl/min for 420 s followed by a 7200-s dissociation over 2-fold increases in concentration between 1.5625 nM and 50 nM. The chip surface was regenerated before each run with a single 20-s injection of 10 mM glycine, pH 1.7 at 30 μl/min where baseline was achieved. The buffer difference was subtracted from blank cell lanes that had been activated and deactivated using standard amine coupling methods. Each concentration series was performed in duplicate. Data generated were processed using the BIAevaluation software (GE Healthcare), where multicycle kinetics were obtained using a 1:1 Langmuir binding model.

Protein Expression and Purification—Recombinant PfRh4 was previously expressed using expression vector pET-45b(+) as described (17, 31, 33). For this paper we digested this clone using BamHI and XhoI and cloned the DNA-encoding PfRh4 in-frame into a pProEX HTa vector, which contains a hexa-His tag and tobacco etch virus cleavage site at the N-terminal end. The fusion protein was expressed in BL21 (DE3) bacteria cells and purified over a nickel-nitrilotriacetic acid column (Qiagen) under native conditions. The purified protein underwent tobacco etch virus protease cleavage overnight at 4 °C to remove 23 amino acids including the hexa-His tag. Cleaved rPfRh4 was further purified on a Superdex 200 gel filtration column (Hiload 16/60; GE Healthcare), and appropriate fractions concentrated and loaded onto a cation exchange column (Mono S 5/50 GL; GE Healthcare). Recombinant rPfRh4 eluted from the column as a monomer and purity was determined using SDS-PAGE.

Protein constructs CCPs 1 and 2 and CCPs 1–3 were expressed and purified as described before (33). In brief, existent *Pichia pastoris* clones were fermented as described in Schmidt *et al.* (38), and the recombinant proteins were secreted into the supernatant and captured by ion chromatography with Sepharose beads. The capture step was followed by cation and size exclusion chromatography. The coding DNA for CCPs 8 and 9 and the mutant versions of CCPs 1 and 2, *i.e. m6–9*, *m18*, and *m20*, as well as the revertant mutant CCPs 8 and 9 *m7–9* and *18–20r*, were codon-optimized, gene-synthesized (IDT Technologies), and subcloned into the *P. pastoris* expression vector pPICZαB. Within CCPs 1 and 2 and its mutant versions the *N*-glycosylation sites were removed by substituting relevant Asn to Gln residues. The novel expression cassettes were transformed into the *P. pastoris* strain KM71H according to the manufacturer's instructions (Life Technologies). Protein expression and purification was carried out as for CCPs 1 and 2 and yielded between 10 and 52 mg of a highly pure protein preparation per liter fermenter supernatant. *N*-Glycosylation (CCP 8 and 9 constructs) was removed with the endoglycosidase $EndoH_f$ (New England BioLabs) when applicable in between purification steps using a published procedure (38). All CCP 1 and 2 and CCP 8 and 9 constructs were submitted to mass spectrometry analysis. Purified proteins were analyzed on an LTQ-Orbitrap Velos Pro (Thermo Scientific) online coupled

to an RSLCnano uPLC (Thermo Scientific) equipped with a Vydac MS C4 column (300 Å, 5 μm, 5 mm × 300 μm) (Grace). After desalting the samples at 5% acetonitrile, 0.1% formic acid with a flow of 10 μl/min for 10 min, proteins were eluted from the column by raising the concentration of acetonitrile to 43% for 5 min. MS spectra were acquired in the Orbitrap part of the instrument at a resolution of 30,000. The acquired data were processed using the QualBrowser (Thermo Scientific) in combination with in-house written VBA-script for MSExcel (Microsoft Corp.). These analyses confirmed the identity of the protein preparations with a maximal deviation from the theoretical molecular mass of 12.5 ppm (or 0.18 Da). Additionally, one-dimensional NMR spectroscopy shows that line widths and signal dispersion (measured at 30 °C in PBS with addition of 7% D_2O) are consistent with mono-dispersed, well folded protein molecules.

Surface Plasmon Resonance Measurement of CCP Mutants and rPfRh4—To analyze the interactions between rPfRh4 with recombinant CR1 constructs, SPR experiments were carried out at 25 °C using a Reichert SR7500DC SPR instrument (Reichert Technologies, Buffalo, New York). Recombinant Rh4 was covalently immobilized onto one flow cell of a carboxymethyldextran hydrogel biosensor chip (CMD500m, purchased from XanTec bioanalytics GmbH, Dusseldorf, Germany) by standard amine coupling according to the manufacturer's instructions. Amine coupling in the absence of any protein (dummy coupling) was performed on a second flow cell yielding a reference flow cell. Signals obtained for the rPfRh4 surface were subtracted by signals obtained for the reference flow cell according to standard procedure. Only reference-subtracted sensorgrams are shown throughout. As running buffer 10 mM HEPES, pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$, and 0.005% Tween 20 was used throughout, apart from one protein series of CCPs 1 and 2 (indicated), which was assayed in the same buffer but with $MgCl_2$ substituted with 3 mM EDTA. Proteins were diluted into running buffer before assaying and were injected on the chip for 3 min at a flow rate of 25 μl/min. The dissociation phase consisted of buffer flow at 25 μl/min for 5 min and was followed by a regeneration step with 1 M NaCl. CCPs 1 and 2, CCPs 1 and 2 *m6–9*, CCPs 1 and 2 *m18*, CCPs 8 and 9 *m7–9,18–20r* (CCPs 8 and 9 *Rev*) and CCPs 8 and 9 were assayed at 1:1 dilution series from 10 μM to 0.61 nM. CCP 1 and 2 *m20* and CCP 8 and 9 were assayed at 1:1 dilution series from 10 μM to 9.8 nM. Affinity constants were extracted by plotting the response at steady state against the molar concentration and subsequent fitting the affinity with the TRACEDRAWER software using a 1:1 steady state affinity model. To probe reproducibility the binding of double injections (at same concentration) were probed for all analytes at the highest concentration assayed (apart from CCP 1 and 2 *m18* and CCPs 1–3) and at the lowest concentration assayed (for all).

Results

Characterization of Anti-PfRh4 Monoclonal Antibodies—PfRh4 binds to CR1 to mediate entry of *P. falciparum* parasites into red blood cells (17). We previously generated a recombinant fragment of PfRh4 (rPfRh4) that binds to red blood cells and forms a complex with sCR1 and CCPs 1–3 (17, 31). We immunized mice with rPfRh4, generated 10 murine mAbs that could detect rPfRh4 by ELISA (Fig. 1*A*), and proceeded to examine whether they bound native PfRH4 or interfered with receptor engagement. 2E8 mAb is specific to the C-terminal end of native PfRh4, which does not include the region encompassed by rPfRh4 and as expected did not show significant reactivity to rPfRh4 (Fig. 1*A*) (17).

PfRh4 together with the other PfRh family of proteins localize to the rhoptries of *P. falciparum* parasites (6). Using an indirect immunofluorescence assay we showed that all mAbs except 6F12 detected PfRh4 in an apical localization that predominantly co-localized with PfRh2, another rhoptry protein (Fig. 1*B*) (39). We tested the ability of the mAbs to detect unfolded PfRh4 from solubilized *P. falciparum* schizont lysate by Western blotting. Native PfRh4 migrates as a doublet at 180 and 190 kDa (31), which was detected by all the mAbs except for 5H12 in a *P. falciparum* strain expressing PfRh4 (Fig. 1*C*). This PfRh4-specific doublet is not present in parasite lysates harvested from a PfRh4 knock-out strain. We hypothesize that 5H12 recognizes a conformational epitope that is not present under SDS-PAGE conditions and which denatures PfRh4.[3] To examine if these mAbs are able to immunoprecipitate folded rPfRh4, we incubated rPfRh4 with mAbs and used protein G-Sepharose beads to capture the antigen-antibody complexes. Only 5H12, 6A5, and 10C9 were able to immunoprecipitate rPfRh4 as seen by its presence in the eluate lane (Fig. 1, *D* and *E lanes*). We also performed similar assays using parasite culture supernatants, which contain a processed 160-kDa fragment of PfRh4 (31), and were able to show that 2E8, 5H12, 6A5, and 10C9 were able to immunoprecipitate PfRh4 (Fig. 1*E*). Collectively these results show that all anti-PfRh4 mAbs have specific reactivity to both rPfRh4 and native PfRh4 from *P. falciparum* parasites. In addition, 5H12 may recognize a conformational epitope on PfRh4.

Modulation of Recombinant PfRh4 and CR1 Interaction with Monoclonal Antibodies—To determine whether anti-PfRh4 mAbs inhibited the interaction between PfRh4 and CR1, we developed a FRET assay where 125 nM rPfRh4 labeled with DyLight-594 could be shown to bind to the same concentration of recombinant CCPs 1 and 2 labeled with DyLight-488 (Fig. 2, *A–D*, *No inh columns*). We validated this assay by introducing 50-fold molar excess of unlabeled rPfRh4 and CCPs 1 and 2 to compete with the FRET pair and found a marked decrease in FRET ratio (Fig. 2*D*, unRh4 and CCPs 1 and 2, respectively). Denaturants such as 3 M guanidinium chloride and 1% sodium dodecyl sulfate also reduced FRET to similar levels. In contrast, proteins that were unable to bind PfRh4 such as CCPs 8 and 9 and bovine serum albumin had no effect on FRET (Fig. 2*D*) (33). For the following FRET assays, we arbitrarily defined 100% binding as the level of binding measured when no inhibitor was added and 0% binding as the level of binding measured in denaturing concentrations of sodium dodecyl sulfate (Fig. 2*D*, *dotted lines*). This assay allowed rapid screening of the mAbs and identified 5H12 mAb as an inhibitor of the rPfRh4-CR1 interaction (Fig. 2*E*).

[3] 5H12 mAb did not detect native PfRh4 by Western blotting in both reducing and non-reducing conditions.

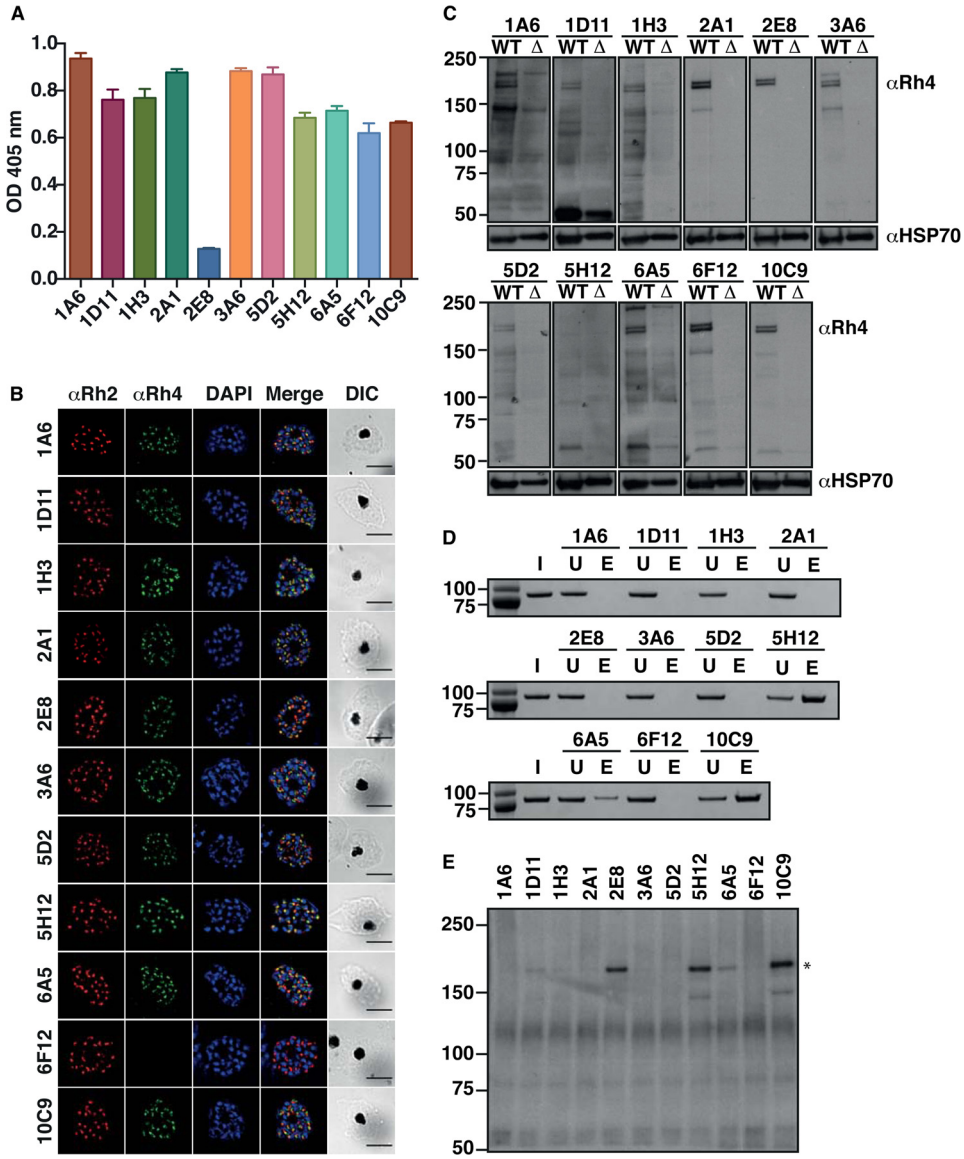


FIGURE 1. **Characterization of monoclonal antibodies specific to PfRh4.** *A*, detection of rPfRh4 by anti-PfRh4 mAbs by ELISA. Anti-PfRh4 mAbs were added to microtiter wells coated with rPfRh4. Bound antigen-antibody complexes were detected with anti-mouse HRP. *Error bars* represent the mean ± S.E. of three independent assays repeated in triplicate. *B*, localization of native PfRh4 was assessed by wide-field immunofluorescence assay using the anti-PfRh4 mAbs. PfRh4 (*green*) was co-stained with rhoptry protein, PfRh2 (*red*), and nuclei-stain DAPI (*blue*). Differential interference contrast (*DIC*) shows the differential interference contrast view of the same field. *Scale bar* = 5 μm. *C*, native PfRh4 is detected by all anti-PfRh4 mAbs except for 5H12. Western blot of saponin-lysed schizont pellet lysates from wild type and ΔPfRh4 strains were separated by SDS-PAGE and probed with anti-PfRh4 mAbs under reducing conditions. *D*, immunoprecipitation of rPfRh4 using anti-PfRh4 mAbs. Recombinant PfRh4 was incubated with anti-PfRh4 mAbs and protein G-Sepharose. Protein eluates were fractionated on SDS-PAGE and visualized using SimplyBlue SafeStain. *I*, input. *U*, unbound. *E*, eluate. *E*, immunoprecipitation of PfRh4 from culture supernatants using anti-PfRh4 mAbs. Culture supernatants were incubated with anti-PfRh4 mAbs and protein G-Sepharose. Protein eluates were fractionated on SDS-PAGE and probed with anti-PfRh4 polyclonal antibody. The *asterisk* marks the processed fragment of PfRh4 present in culture supernatants. For all panels, molecular weight is indicated on the *left* in kDa.

To confirm the results obtained from the FRET-based assay, we examined the effect of mAbs on PfRh4-CR1 interaction by ELISA (Fig. 2*F*). Recombinant PfRh4 was incubated briefly with each mAb before binding to immobilized CCPs 1–3, and bound PfRh4 was detected using a polyclonal anti-PfRh4 antibody. The addition of CCPs 1–3 as a known inhibitor showed the expected marked reduction in PfRh4-CR1 binding (Fig. 2*F*) (33). 5H12 showed similar levels of inhibition of the PfRh4-CR1 interaction as in the FRET assay. We observed that two mAbs, 6A5 and 10C9, appeared to enhance the interaction between rPfRh4 and CCPs 1–3 (Fig. 2*F*).

We examined whether the inhibitory mAb 5H12 and the enhancing mAb 10C9 were able to co-immunoprecipitate PfRh4 and CCPs 1–3. Both 5H12 and 10C9 immunoprecipitate rPfRh4 with similar efficiencies as indicated by the depletion of rPfRh4 in the unbound fraction (*lane U*) and an enrichment of rPfRh4 in the eluate fraction (Fig. 2*G*, *left panel*, *lane E*). When CCPs 1–3 was added into the assay, 10C9 was able to immunoprecipitate the rPfRh4-CCPs 1–3 complex (Fig. 2*G*, *right panel*). In contrast, although 5H12 successfully immunoprecipitated rPfRh4, we did not detect a corresponding CCPs 1–3 protein (Fig. 2*G*, *right panel*). 5H12 was less efficient at

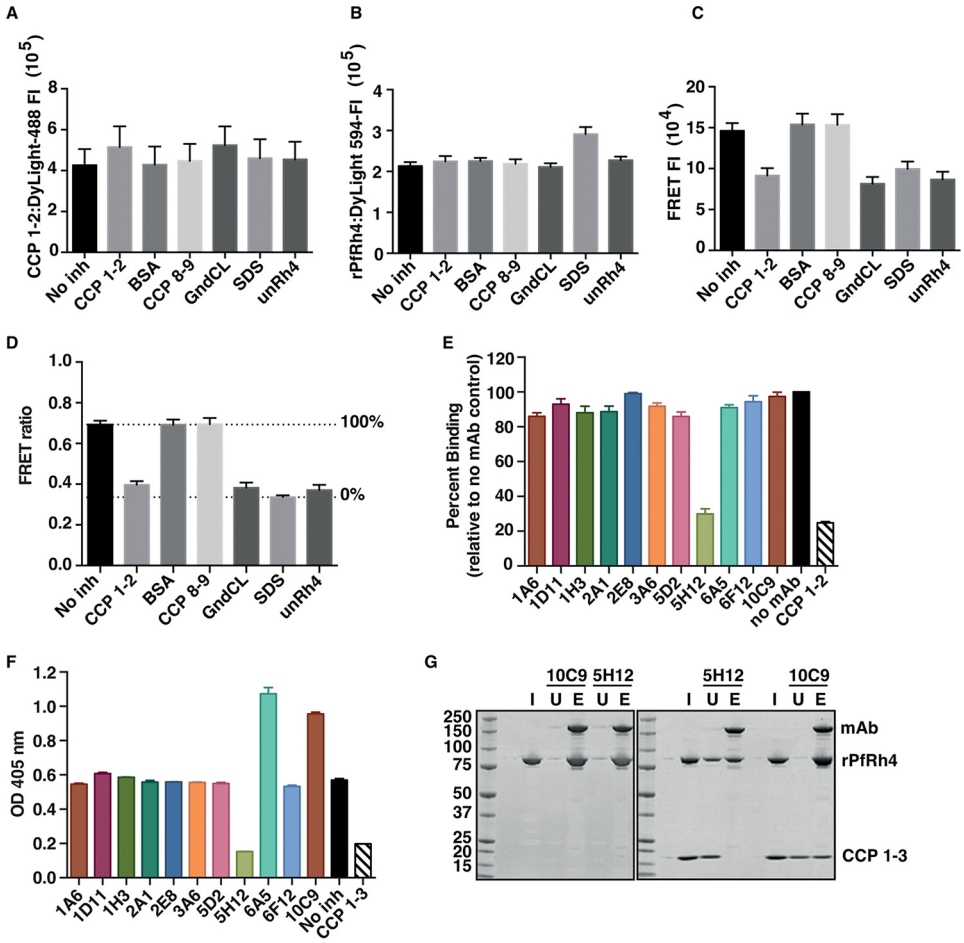


FIGURE 2. **Inhibition of the PfRh4-CR1 interaction with anti-PfRh4 mAbs.** Novel FRET-based assay to monitor the interaction between PfRh4 and CR1 where CCPs 1 and 2 (*A*) and rPfRh4 (*B*) were labeled with Dylight 488 and 594, respectively, and incubated at 1:1 molar concentration (*no inh*) either in the presence of unlabeled proteins (CCPs 1 and 2, BSA, CCPs 8 and 9, *unRh4*) or denaturants (guanidine hydrochloride (*GndCL*) and SDS). The fluorescence intensity (*FI*) of DyLight-488 (donor) was measured with a 485/14-nm excitation filter and a 535/25-nm emission filter and DyLight-594 was measured with a 590/20-nm excitation filter and 615/9-nm emission filter. *C*, measurement of FRET fluorescence intensity using the DyLight-488 (donor) 485/14-nm excitation filter and the DyLight 594 (acceptor) 615/9-nm emission filter. *D*, FRET ratio on the *y* axis represents "excitation 488/emission 615" value over Dylight 594 value. Unlabeled CCPs 1 and 2 and rPfRh4 (*unRh4*) were able to inhibit labeled PfRh4-CR1 interaction on a similar level to denaturants, SDS and guanidine hydrochloride (GndCL). *Error bars* represent the S.E. of three independent repeats. *E*, 5H12 inhibits the PfRh4-CR1 interaction in the FRET-based assay. Dylight-labeled rPfRh4 and CCPs 1 and 2 were incubated with various anti-PfRh4 mAbs. Percentage binding refers to the FRET ratio relative to "no mAb" and SDS controls. *Error bars* represent S.E. of three independent repeats. *F*, 5H12 inhibits PfRh4-CR1 interaction in ELISA-based assay. Microtiter wells were coated with CCPs 1–3. Recombinant PfRh4 was added in the presence of anti-Rh4 mAbs. Bound rPfRh4 was detected with an anti-PfRh4 rabbit polyclonal primary antibody followed by an anti-rabbit-HRP secondary antibody. *Error bars* represent S.E. of two independent repeats. *G*, 5H12 disrupts PfRh4-CR1 complex formation. The immunoprecipitation assay was performed by incubating rPfRh4 with 5H12 and 10C9 (*left panel*) and subsequently in conjunction with CCPs 1–3 (*right panel*). *I*, input. *U*, unbound. *E*, eluate.

immunoprecipitating rPfRh4 when CCPs 1–3 were present, suggesting that 5H12 binds to a region of rPfRh4 that is also bound by CCPs 1–3 (Fig. 2*G*, *right panel versus left panel*).

Blocking the Red Blood Cell Binding Capabilities of PfRh4—We utilized a flow cytometry-based red blood cell binding assay to determine the effects of the mAbs on PfRh4 interaction with native CR1 on the surface of red blood cells. This method entailed incubating rPfRh4 with red blood cells in the presence of mAbs; rPfRh4 binding was detected using an anti-PfRh4 rabbit polyclonal antibody followed by an anti-rabbit secondary antibody conjugated to a fluorophore, which was detected by flow cytometry. This assay was used previously to show that the levels of rPfRh4 binding to erythrocytes directly correlates with the levels of CR1 on the red blood cell surface (17). For this assay we used an intermediate amount of rPfRh4 that allows both increases and decreases in rPfRh4 binding to be measured. As expected, the addition of soluble CCPs 1 and 2 was able to inhibit rPfRh4 binding to red blood cells (Fig. 3, *A* and *B*). 1D11, 5H12, 6A5, and 10C9 were all able to affect PfRh4 binding to red blood cells. Consistent with previous results, 5H12 was able to inhibit rPfRh4 binding to a similar extent as CCPs 1 and 2 (Fig. 3, *A* and *B*). On the other hand, 1D11 and 10C9 showed strong enhancement of rPfRh4 erythrocyte binding, whereas 6A5 increased rPfRh4 binding more modestly (3.6, 19.5, and 1.5-fold increase, respectively).

We tested if the mAbs could inhibit the PfRh4-CR1 invasion pathway in *P. falciparum* growth assays. As a negative control in these experiments we used the 2E8 mAb, which is specific to the C-terminal end of native PfRh4 (not present in rPfRh4) and has previously been shown to be unable to inhibit parasite growth (17). *P. falciparum* parasites invade red blood cells via several redundant invasion pathways mediated by EBA and PfRh proteins. To abolish the pathways mediated by the EBA family of proteins, red blood cells were treated with neuramin-

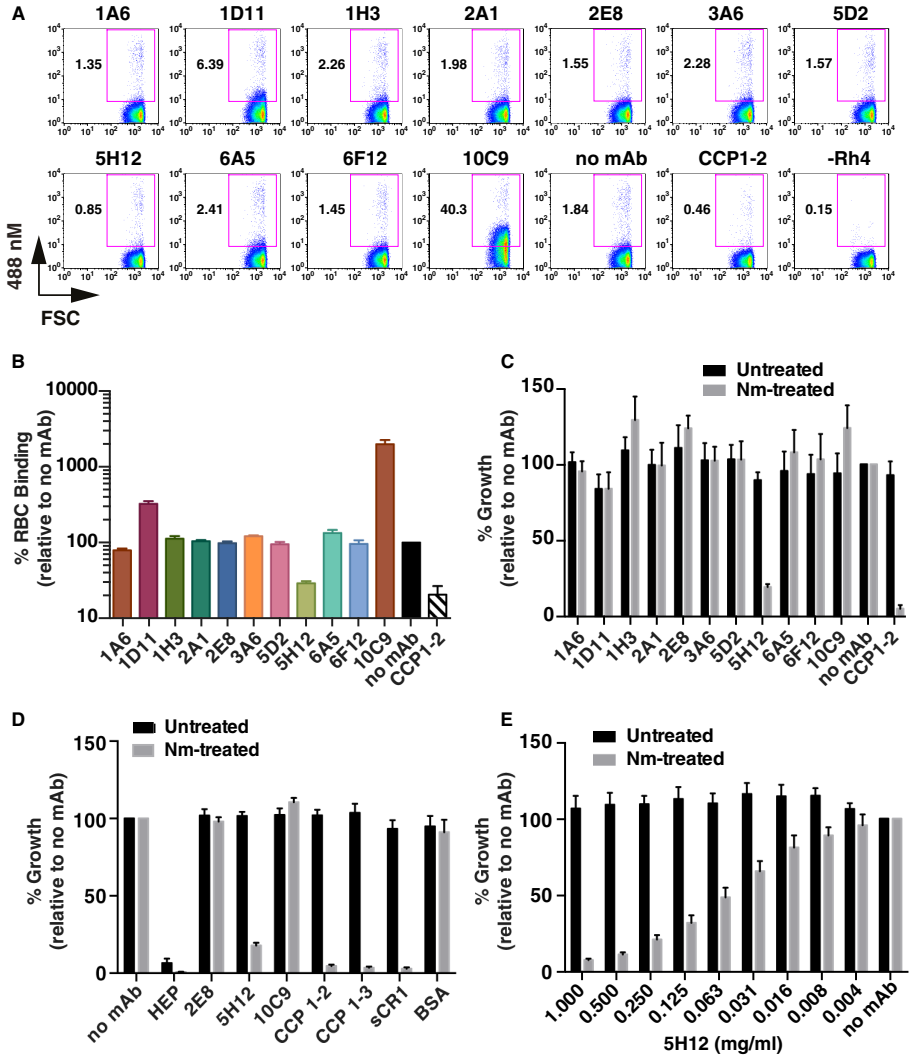


FIGURE 3. **Inhibition of the PfRh4 invasion pathway.** *A*, dot plots show the binding of rPfRh4 to red blood cells in the presence of anti-PfRh4 mAbs. Binding was detected using an anti-PfRh4 rabbit IgG antibody followed by a secondary anti-rabbit Alexa 488 antibody. *Numbers outside of the purple box* refer to the percentage of red blood cells with bound rPfRh4 relative to the full red blood cell population. *B*, recombinant PfRh4 binding to erythroid CR1 was inhibited by 5H12. Binding assays were performed by incubating rPfRh4 and red blood cells in the presence of various anti-PfRh4 mAbs, and the level of red blood cell binding was determined by flow cytometry. Percentage RBC binding on the *y* axis refers to the percentage of rPfRh4-erythroid CR1 binding relative to the *no mAb* control (arbitrarily set to be 100%). *Error bars* represent S.E. of three independent repeats. *C*, anti-PfRh4 mAbs were tested with untreated (*black bars*) or neuraminidase-treated erythrocytes (*gray bars*) at a final concentration of 0.2 mg/ml. Growth inhibition assays were performed using the D10-PHG strain of *P. falciparum*. *D*, anti-PfRh4 mAb, 5H12, inhibits the PfRh4-CR1 pathway to similar levels as the recombinant CR1 constructs. D10-PHG strain was tested in untreated (*black bar*) or neuraminidase-treated erythrocytes (*gray bar*) in the presence of anti-PfRh4 mAbs (at a final concentration of 0.2 mg/ml) and recombinant CR1 constructs (at a final concentration of 0.1 mg/ml). *Error bars* represent S.E. from five independent repeats. *E*, titration of 5H12 in growth inhibition assays in the presence of untreated (*black bar*) or neuraminidase-treated red blood cells (*gray bar*). *Error bars* represent S.E. from four independent repeats. For *panels C*, *D*, and *E*, percent growth on the *y* axis refers to the percentage parasitemia relative to the no mAb control, which is arbitrarily set at 100%.

idase to remove sialic acid residues from the glycophorins, which are cognate receptors for EBA proteins. Upon neuraminidase treatment, 90% of the invasion events are dependent on PfRh4 for entry into red blood cells via CR1 (17, 33). When all 10 mAbs and 2E8 were tested in neuraminidase-treated red blood cells, we observed inhibition of the PfRh4-CR1 pathway only when 5H12 was added to the growth medium (Fig. 3*C*). In subsequent growth assays we additionally tested CR1 fragments, 5H12, and the enhancing mAb 10C9. The addition of CCPs 1 and 2, CCPs 1–3, sCR1 inhibited >90% of the invasion events in neuraminidase-treated red blood cells as expected (Fig. 3, *C* and *D*), whereas the addition of 2E8 and 10C9 mAb did not perturb growth in a statistically significant manner. 5H12 again exhibited similar levels of growth inhibition as soluble forms of CR1 (Fig. 3*D*). Titration of 5H12 showed dose-dependent inhibition of *P. falciparum* invasion events and growth in neuraminidase-treated red blood cells (Fig. 3*E*).

Mapping the Epitopes for Monoclonal Antibody Binding—Having characterized the functional outcomes of mAb binding to PfRh4, we next utilized a panel of overlapping 15-mer peptides completely encompassing rPfRh4 to identify epitopes recognized by the anti-PfRh4 mAbs (Fig. 4). Using ELISA we detected specific binding to either a single peptide or a set of overlapping peptides for all of the mAbs except 1D11, 1H3, and 5H12. Both 1A6 and 3A6 bound two overlapping peptides: N238 (NEKLEKYTNKFEHNI) and E242 (EKYTNKFEHNIKPHI). 5D2 bound a single peptide (Y266, YINNSDCHLTCSKYK). 2A1 bound three overlapping peptides Y522 (YDNIYIIL-

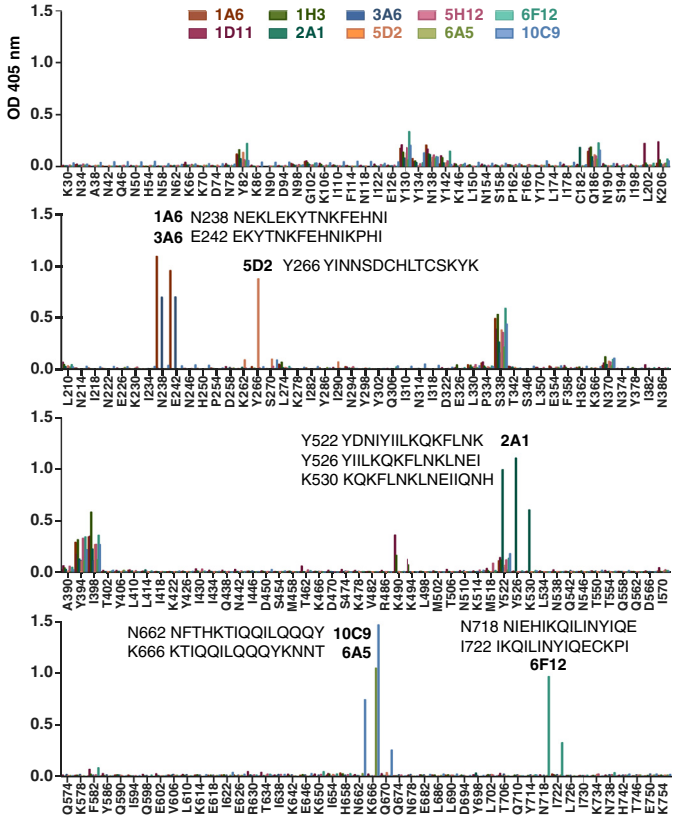


FIGURE 4. **Epitope mapping of PfRh4 monoclonal antibodies.** Detection of anti-PfRh4 mAbs binding to an overlapping peptide array encompassing rPfRh4 by ELISA. Peptides are named by the single-letter code and numerical position of their first amino acid within PfRh4; *e.g. K30* indicates the peptide KEKKNDPEADSKNSQ starting at amino acid position 30. Significant binding of anti-PfRh4 mAbs to peptides is highlighted with the peptide sequence.

KQKFLNK), Y526 (YIILKQKFLNKLNEI), and K530 (KQKFLNKLNEIIQNH). 10C9 bound two overlapping peptides, N662 (NFTHKTIQQILQQQY) and K666 (KTIQQILQQQYKNNT), whereas 6A5 also recognized peptide K666. It is interesting to note that both 10C9 and 6A5 enhance rPfRh4 red blood cell binding. 6F12 bound to novel epitopes encompassed by two overlapping peptides, N718 (NIEHIKQILINYIQE) and I722 (IKQILINYIQECKPI).

To determine if any of the mAbs share a similar epitope to 5H12, we also performed competition ELISA between 5H12 (which was directly conjugated to HRP, 5H12-HRP) and all the other mAbs. Using rPfRh4-coated microtiter plates, we examined the level of 5H12-HRP binding in the presence of each mAb by ELISA. Although the addition of 5H12 itself successfully competed with 5H12-HRP, none of the other mAbs showed a similar phenotype (data not shown).[4] These data show that the 5H12 epitope is unique and does not overlap with other epitopes bound by the collection of anti-PfRh4 mAbs.

Antibody Affinities for PfRh4—The affinity of two mAbs, 5H12 and 10C9, to rPfRh4 was evaluated using SPR. 5H12 and 10C9 showed clear binding (Fig. 5, *A* and *B*, respectively). rPfRh4 bound to 5H12 with low nanomolar affinity (K_D = 1.34 nM) and with association (k_a = 1.29 × 10^5 $M^{-1}s^{-1}$) and dissociation (k_d = 1.73 × 10^{-4} s^{-1}) (Table 1). In comparison, 10C9 showed an even higher affinity for rPfRh4 (K_D = 0.48 nM), with association (k_a = 5.64 × 10^4 $M^{-1}s^{-1}$) and dissociation (k_d = 3.69 × 10^{-5} s^{-1}) (Table 1).

CR1-based Inhibitors That Block PfRh4-CR1 Interaction—Mutagenesis analyses within CCPs 1–7 identified several residues essential for PfRh4 binding within CCP 1 (Fig. 6*A*) (32). Although these mutations clearly abolished PfRh4-CR1 interaction *in vitro* as shown by ELISA, immunoprecipitation, and SPR measurements, the mutant proteins were not assessed for their functional effects in red blood cell binding or growth assays. To this end we generated recombinant CCPs 1 and 2 proteins in *P. pastoris* that contained mutations of these critical residues (Fig. 6*B*). These mutations (m) are named in relation to their positions in CCP 1 that were substituted with the respective amino acids in CCP 8 (Fig. 6*C*) (32). We produced correctly folded CCPs 1 and 2 *m6–9*, CCPs 1 and 2 *m18*, and CCPs 1 and 2 *m20*. In addition, we also expressed an engineered PfRh4 binding site in CCPs 8 and 9 by introducing CCP 1 residues at 7–9 and 18–20 to their homologous site within CCP 8 (32). This mutant is called CCPs 8 and 9 *m7–9,18–20r* but will be referred to as CCPs 8 and 9 *Rev* for simplicity. The proteins were purified to a high degree of purity, as judged by prominent single protein bands under reducing and non-reducing conditions after SDS-PAGE analysis (Fig. 6*B*). The faster mobility under non-reducing conditions is indicative of disulfide bridges being formed within these constructs. Protein identity and quality have additionally been confirmed by mass spectrometry with a maximal deviation of 12.5 ppm (equating to 0.18 Da) from the theoretical molecular mass.

Using immunoprecipitation assays, we examined whether these mutants could still form a complex with rPfRh4. We observed that CCPs 1 and 2 was able to form a stable complex with rPfRh4 (Fig. 6, *D* and *E*). Unexpectedly CCPs 1 and 2 *m6–9* and CCPs 1 and 2 *m18* were still able to interact with rPfRh4 (32). Mutation of residue 20 (CCPs 1 and 2 *m20*) completely abolished complex formation. As expected, CCPs 8 and 9 did not interact with rPfRh4, whereas CCPs 8 and 9 *Rev* was able to form a stable complex with rPfRh4 (Fig. 6*D*, *right panel*), indicating that the substituted residues from CCP 1 to CCP 8 are sufficient to mediate PfRh4 binding.

To determine if CCPs 8 and 9 *Rev* bound to PfRh4 using the same binding site as recognized by the inhibitory antibody, we examined if the addition of mAb 5H12 would interfere with the rPfRh4-CCPs 8 and 9 *Rev* complex. 10C9 was able to immunoprecipitate rPfRh4-CCPs 8 and 9 *Rev* complex (Fig. 6*E*). However, 5H12 successfully immunoprecipitated rPfRh4, but we did not detect a corresponding CCPs 8 and 9 *Rev* fragment (Fig. 6*E*), suggesting that 5H12 inhibits the interaction between rPfRh4 and CCPs 8 and 9 *Rev*.

We utilized the flow cytometry-based red blood cell binding assay described earlier to determine the effects of the mutant proteins. As expected, the addition of soluble CCPs 1 and 2, CCPs 1 and 2 *m6–9*, CCPs 1 and 2 *m18*, and CCPs 8 and 9 *Rev* was able to inhibit rPfRh4 binding to red blood cells (Fig. 6*F*). Proteins that did not bind rPfRh4 such as CCPs 8 and 9 and CCPs 1 and 2 *m20* were not able to inhibit rPfRh4 binding to red blood cells (Fig. 6*F*). Consistent with these results, we observed

[4] Using competition ELISA, only unlabelled 5H12 mAb could outcompete 5H12-HRP for binding to immobilized rPfRh4.

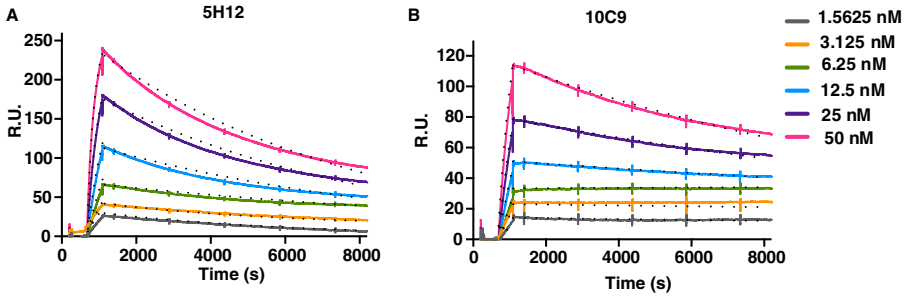


FIGURE 5. **Kinetics of mAb-PfRh4 interactions.** Sensorgrams of multi-cycle kinetic measurements of 5H12 (*A*) and 10C9 (*B*) with PfRh4. The sensorgrams represent duplicate runs obtained on separate days with increasing concentrations of rPfRh4 (1.5625, 3.125, 6.25, 12.5, 25, 50 nM) flowed over 5H12 and 10C9 captured on anti-mouse IgG. *Solid lines* represent the observed binding, and *dots* represent 1:1 Langmuir model-fitted binding with kinetic values reported in Table 1. *R.U.*, resonance units.

TABLE 1

Multicycle kinetic measurements of mAb-PfRh4 interactions

Kinetic parameters (k_a and k_d) and K_D values extracted from 5H12-rPfRh4 and 10C9-rPfRh4 sensogram fitted to a global 1:1 interaction model are reported. Maximum resonance values (R_{max}) and χ^2 values of goodness of fit (<10% R_{max}) for the models are also presented. RU, resonance units.

mAb	k_a	k_d	R_{max}	K_D	χ^2
	$M^{-1}\,s^{-1}$	*1/s*	*RU*	*M*	
5H12	1.29E+05	1.73E−04	236	1.34E−9	2.7
10C9	5.64E+04	3.69E−05	111	4.82E−10	2.3

that the addition of CCPs 1 and 2, CCPs 1 and 2 *m6–9*, CCPs 1 and 2 *m18*, and CCPs 8 and 9 *Rev* was able to inhibit the majority of *P. falciparum* invasion events into neuraminidase-treated red blood cells (Fig. 6*G*). This was similar to the level of inhibition seen upon the addition of CCPs 1–3 and sCR1. As expected, we did not observe any inhibition of the PfRh4-CR1 invasion pathway with the addition of CCPs 8 and 9 or CCP 1 *m20* (Fig. 6*G*).

To ensure that the protein fragments that were used in all the assays above were properly folded, we also performed one-dimensional NMR spectroscopy. One-dimensional NMR spectroscopy shows that line widths and signal dispersion are consistent with monodispersed, well folded protein molecules (Fig. 7, *A–F*).

Affinity of CR1-based Inhibitors for Binding to PfRh4—We expected that mutations in CCPs 1 and 2 and CCPs 8 and 9 would directly influence each mutant protein's affinity for PfRh4. To determine the binding constants of the wild type and mutant proteins, we employed SPR. rPfRh4 was immobilized onto a carboxymethyldextran hydrogel sensorchip surface, and the CR1 constructs were assayed in HEPES-buffered saline containing 1 mM $MgCl_2$ (Fig. 8). CCPs 1 and 2, CCPs 1 and 2 *m6–9*, and the revertant mutant CCPs 8 and 9 *Rev* exhibited the highest affinities for rPfRh4 binding (20.0, 33.5, and 64.9 nM; Fig. 8, *A*, *B*, and *F*, respectively), which is in accordance with their inhibitory potential. CCPs 1 and 2 *m18* interacts with rPfRh4 with an affinity of 336 nM (Fig. 8*C*), which is markedly reduced from the CCPs 1 and 2 binding constant of 20 nM. As expected CCPs 1 and 2 *m20* and CCPs 8 and 9 failed to produce any substantial binding response (Fig. 8, *D* and *E*). Although the determined affinity constants for the CR1-based inhibitors are in excellent correlation with their biological activities in this study, we noted that the binding constants differ from previous studies, in which CCPs 1–7 or CCPs 1–3 were found to interact with PfRh4 with affinities in the range from 0.5 to 11 μM (32, 33). To rule out that the presence of CCP 3 (within the construct CCPs 1–3) alters the binding affinities to rPfRh4, we also performed affinity measurements using CCPs 1–3 (Fig. 8*G*). With a K_D of 39.2 nM CCPs 1–3 exhibit similar binding properties when compared with CCPs 1 and 2, ruling out a prominent influence of CCP 3 on the underlying interaction. Because the previous studies have assayed the binding behavior in HEPES buffer in the absence of $MgCl_2$, we also interrogated if the presence of magnesium ions could have modified the interaction. However, the presence or absence of magnesium ions did not influence the binding of CCPs 1 and 2 to rPfRh4 with the K_D in 1 mM $MgCl_2$ or 3 mM EDTA buffer being fitted to 20.0 or 16.9 nM (Fig. 8, *A* and *H*, respectively). Thus, the markedly tighter association between the CR1 constructs and rPfRh4 observed in this study likely originates from a modified purification protocol of the rPfRh4 construct, which is discussed below. Using SPR, we determined that the inhibitory biological activity for the CCPs 1 and 2 and CCPs 8 and 9 mutant constructs correlates remarkably well with the affinities for their binding partner rPfRh4, which are in the lower nanomolar range.

Discussion

Parasite entry into red blood cells is an essential component of the *P. falciparum* life cycle in humans. *P. falciparum* has evolved to use redundant invasion pathways, and through the ability to switch pathways via differential expression of parasite adhesins, the parasite has more opportunities for successful invasion in the face of the human immune response and red blood cell polymorphisms prevalent in malaria endemic regions. By activating expression of PfRh4, the parasite is able to switch receptor usage from sialic acid-dependent to sialic acid-independent pathways, thus providing a mechanism for the parasite to invade via non-glycophorin mediated entry. In the work reported here we characterize a repertoire of anti-PfRh4 mAbs including the first anti-PfRh4 neutralizing antibody (summarized in Table 2) and describe a collection of CR1-based inhibitors, which are able to block PfRh4 interaction with CR1 and inhibit *P. falciparum* sialic-acid independent invasion pathways. Collectively, these results will guide future work on the development of single epitope inhibitors of the CR1-PfRh4 invasion pathway.

5H12 mAb represents the first anti-PfRh4 neutralizing antibody. Using three robust PfRh4-CR1 interaction assays based on ELISA, FRET, and immunoprecipitation, we show that 5H12

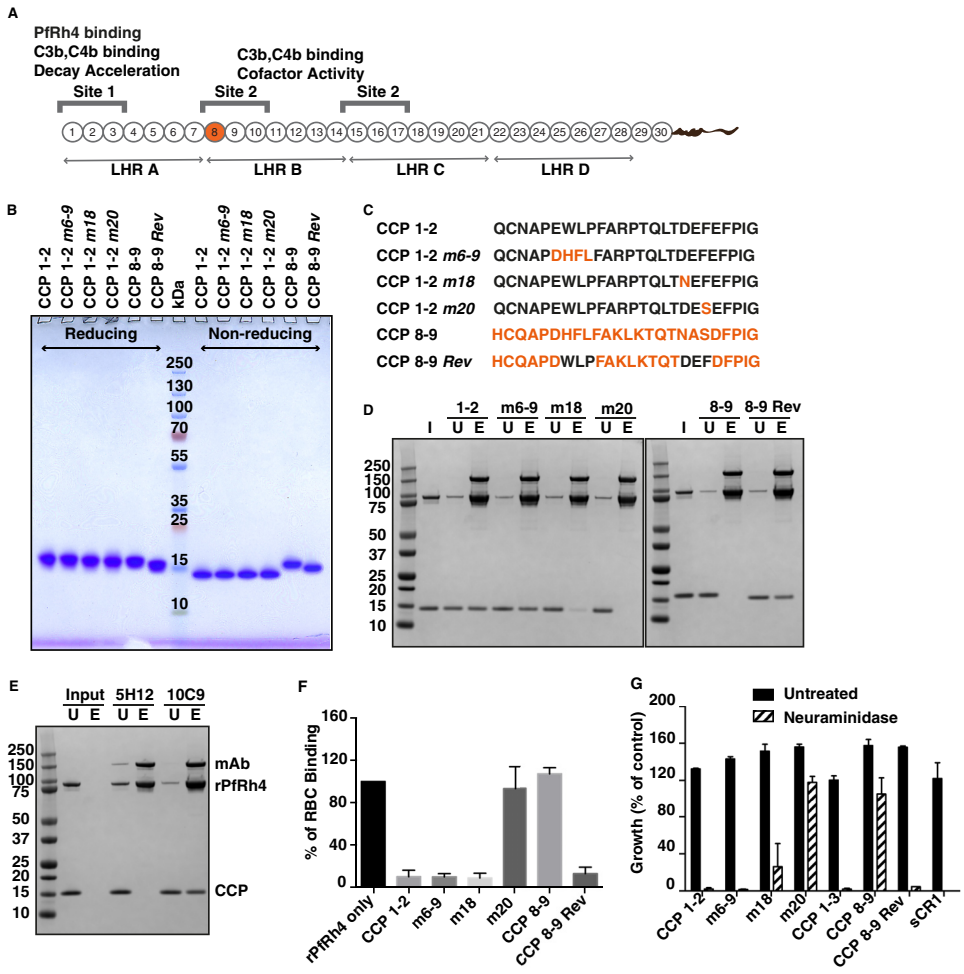


FIGURE 6. **CR1-based inhibitors of PfRh4-CR1 interaction.** *A*, schematic of CR1 polypeptide. Each *box* labeled 1–30 represents a CCP module of 60–70 amino acid residues. The first 28 CCPs are organized based on homology into four long homologous repeats (*LHR*) *A–D*, each consist of seven CCPs. The functional sites of CR1 are labeled *Site 1* and *Site 2* with their respective activities. The transmembrane domain and cytoplasmic tail are represented as a *curvy line at the C-terminal end*. *B*, SDS-PAGE gel of purified CCP constructs. 2 μg of each construct were loaded onto a NuPAGE® gradient gel (4–12%) under reducing and non-reducing conditions and stained with Coomassie Brilliant Blue. Molecular weight markers are indicated in kDa in the *middle of the gel*. *C*, the first 25 amino acids within CCPs 1 and 2 and CCPs 8 and 9 are shown. Wild type CCP 1 and CCP 8 sequence is in *black* and *orange*, respectively. *Numbers* after the *m*, for mutants, denote the position of the amino acids. *D*, immunoprecipitation assay to monitor PfRh4-CR1 complex formation. Anti-PfRh4 antibody 10C9 was incubated with rPfRh4 in the presence of either CCPs 1 and 2, CCPs 1 and 2 *m6–9*, CCPs 1 and 2 *m18*, CCPs 1 and 2 *m20*, CCPs 8 and 9, or CCPs 8 and 9 *Rev*. Proteins were stained with SimplyBlue SafeStain. *I*, input. *U*, unbound. *E*, eluate. *E*, inhibitory antibody 5H12 is able to block complex formation between CCPs 8 and 9 *Rev* and rPfRh4. *F*, effects of mutagenesis within CCPs 1 and 2 on rPfRh4 binding to red blood cells. Percentage RBC binding of rPfRh4 on the *y* axis refers to the percentage of rPfRh4-erythroid CR1 binding relative to the rPfRh4 only control (arbitrarily set to be 100%). *Error bars* represent ± S.E. from three independent repeats. *G*, growth assays in the presence of CR1 recombinant constructs. Various CCP mutants and soluble CR1 were tested with untreated (*black bars*) or neuraminidase-treated erythrocytes (*gray bars*) at a final concentration of 0.1 mg/ml. Growth (percentage of control) on the *y* axis refers to the percentage parasitemia relative to the PBS control. *Error bars* represent S.E. from three independent repeats.

is able to disrupt PfRh4 interaction with CR1. The addition of 5H12 also blocks PfRh4 binding to red blood cells and inhibits the PfRh4-CR1 pathway in *P. falciparum* growth assays. A recent study reporting the atomic resolution structure of PfRh5 in complex with its inhibitory antibodies QA1 and QA5 show that the epitopes for these antibodies overlap with the basigin N- and C-terminal domain binding site, respectively. We hypothesize that 5H12 would bind to a conformation-dependent epitope that overlaps with the CR1 binding site in PfRh4 and in effect blocks the ability for the receptor to bind. Epitope mapping of 5H12 with PfRh4 using x-ray crystallography will allow us to identify the binding interface and guide the design of inhibitory antibody epitopes.

6A5 and 10C9 mAb are able to enhance PfRh4 binding to CR1 on red blood cells. Epitope mapping shows that 10C9 bound two overlapping peptides, N662 (NFTHKTIQQ-ILQQQY) and K666 (KTIQQILQQQYKNNT), whereas 6A5 recognized only peptide K666. The striking feature of these peptides is the presence of polyglutamine repeats. As these epitopes are present outside of the minimal red blood cell binding domain of PfRh4 (31), the binding of the mAbs may elicit an allosteric conformational change that enhances the binding capabilities of PfRh4 to red blood cells. Previously we observed that the addition of rabbit polyclonal antibodies raised to the C-terminal end of native PfRh4 were able to enhance the binding of PfRh4 to red blood cells (data not shown).[5] Examination

[5] The addition of rabbit polyclonal antibodies raised to the C-terminal end of PfRh4 was able to enhance binding of native PfRh4 to red blood cells with no effect on EBA-175 binding.

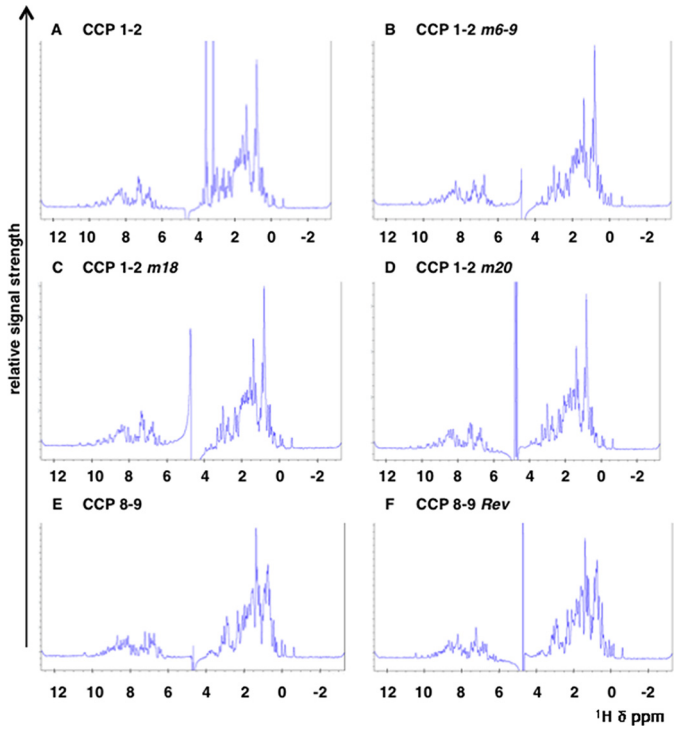


FIGURE 7. **^{1}H NMR spectra of CR1 protein fragments.** Spectra for the CR1 constructs CCPs 1 and 2 (*A*), CCPs 1 and 2 *m6–9* (*B*), CCPs 1 and 2 *m18* (*C*), CCPs 1 and 2 *m20* (*D*), CCPs 8 and 9 (*E*), and CCPs 8 and 9 *Rev* (*F*). The spectra were recorded on a 800-MHz spectrometer from Bruker at 303 K in PBS with the addition of D_2O to 7%. All samples were measured at a protein concentration of 50 μM. The dispersion and line shape as well as peak intensities are comparable among the different samples and indicate comparable high quality of the sample preparations. The overall good signal dispersion and the sharp line widths are consistent with mono-dispersed, well folded protein molecules. The spectrum of CCPs 1 and 2 (*A*) exhibits two strong signals between 3 and 4 ppm. These signals relate to a small contamination with glycerol that derives from the protein concentration process using a commercial spin concentrator (which typically use glycerol to preserve the filtration membrane of the spin concentrator). The signal at 4.8 ppm relates to residual water signal.

of the protein sequence highlighted the presence of QQ*XXXX*-QQQ and a triple QQQ site within the recombinant protein used as immunogen. Antibodies that bind to these sites may elicit an allosteric change in full-length PfRh4, resulting in enhanced binding to CR1, although the molecular mechanism remains to be experimentally verified. It will be important to exclude the epitopes above in future immunizations as not to generate antibodies that enhance the interaction between PfRh4 and CR1.

We also describe four CR1-based inhibitors; CCPs 1 and 2, CCPs 1 and 2 *m6–9*, CCPs 1 and 2 *m18*, and CCPs 8 and 9 *Rev*. These inhibitors competed for PfRh4 binding to erythroid CR1 and block the use of the PfRh4-CR1 invasion pathways in *P. falciparum*. Their inhibitory effects directly correlate with their affinity for PfRh4. The least potent inhibitor CCPs 1 and 2 *m18* has a K_D of 336 nM compared with the 33, 65, and 20 nM for CCPs 1 and 2 *m6–9*, CCPs 8 and 9 *Rev*, and CCPs 1 and 2, respectively. In contrast to the affinity constants for the rPfRh4-CR1 interaction being in the lower nanomolar range as determined in this study, two previous studies have found the binding constants for the interaction of CR1 CCPs 1–7 or CCPs 1–3 to be in the range between 0.5 and 11 μM (32, 33). This former work has also employed SPR to measure the affinity constants for the binding of sensor chip-immobilized rPfRh4 to CR1 CCPs 1–3 or CCP s1–7. Although the CCPs 1–3 construct was overexpressed by *P. pastoris* and purified to homogeneity, the CCPs 1–7 wild type and mutant constructs were expressed by HEK293T with the proteins being directly harvested from serum-free supernatants or captured by metal affinity chromatograph facilitated by a hexa-His tag. Despite differences in glycosylation status and purity, the CCPs 1–3 and CCPs 1–7 preparations exhibited an overall similar binding behavior for rPfRh4 (0.5–11 μM).

In our present study we measured the affinity of *Pichia*-produced CCPs 1–3 for rPfRh4 alongside the novel wild type and mutant constructs within CCPs 1 and 2 and observed a markedly tighter interaction than described before (32, 33). Because the CCP 1 and 3 proteins in all studies have been prepared by the same procedures, a possible reason for this remarkable shift toward higher affinity likely originates in the modification of the purification protocol for rPfRh4, which now includes tobacco etch virus cleavage of the N-terminal hexa-His tag followed by ion exchange chromatography. The fact that the removal of 23 amino acids including the purification tag leads to a substantial increase in affinity argues that accessibility of the N terminus in PfRh4 is critical for its interaction with CR1. In terms of specific amino acids within CCP 1 that are critical for binding PfRh4, previous work identified that *m6–9* exhibited a >60% loss of affinity for rPfRh4, whereas *m18* and *m20* mutants completely abolished the interaction *in vitro*. Although we are able to establish that *m20* is a critical residue involved in mediating PfRh4 interaction, we observed that *m6–9* and *m18*, although lowering the affinity to various degrees, were not essential. We do note, however, that *m18* did result in a strong reduction in PfRh4 binding, whereas the effect of the *m6–9* substitution is negligible. Possible reasons for the different outcomes may originate from the different preparations of rPfRh4, resulting in the presence or absence of the purification tag.

Structurally, our study refines the previously proposed binding site and suggests that the surface-exposed, hydrophobic residue Phe-20 plays the primary role in binding PfRh4 with some contribution from the negatively charged side chain of Asp-18, which on the surface of CCP 1 lies directly next to the aromatic side chain of Phe-20.

Park *et al.* (32) were able to engineer an artificial binding site within CCPs 8–14 by substituting residues within CCP 1 that are critical for PfRh4 interaction to their homologous position in CCP 8. This artificial binding site carrying six CCP 1 amino acid substitutions, *m7–9,18–20r*, bound ~30-fold better than the wild type binding site as measured by SPR. Consistent with this, our results show that CCPs 8 and 9 *Rev* was able to form a stable complex with rPfRh4, whereas CCPs 8 and 9 did not. With 20 and 65 nM we found the affinities between rPfRh4 and CCPs 1 and 2 or CCPs 8 and 9 *Rev*, respectively, to be in a similar range. This observation is in agreement with the comparable biological activity observed for these two constructs. Although we also observed conversion of CCPs 8 and 9 into an active PfRh4 binding partner upon homologous substitution of six amino acids, we were unable to confirm that CCPs 8 and 9 *Rev* bound substantially tighter to PfRh4 than CCPs 1 and 2. A pos-

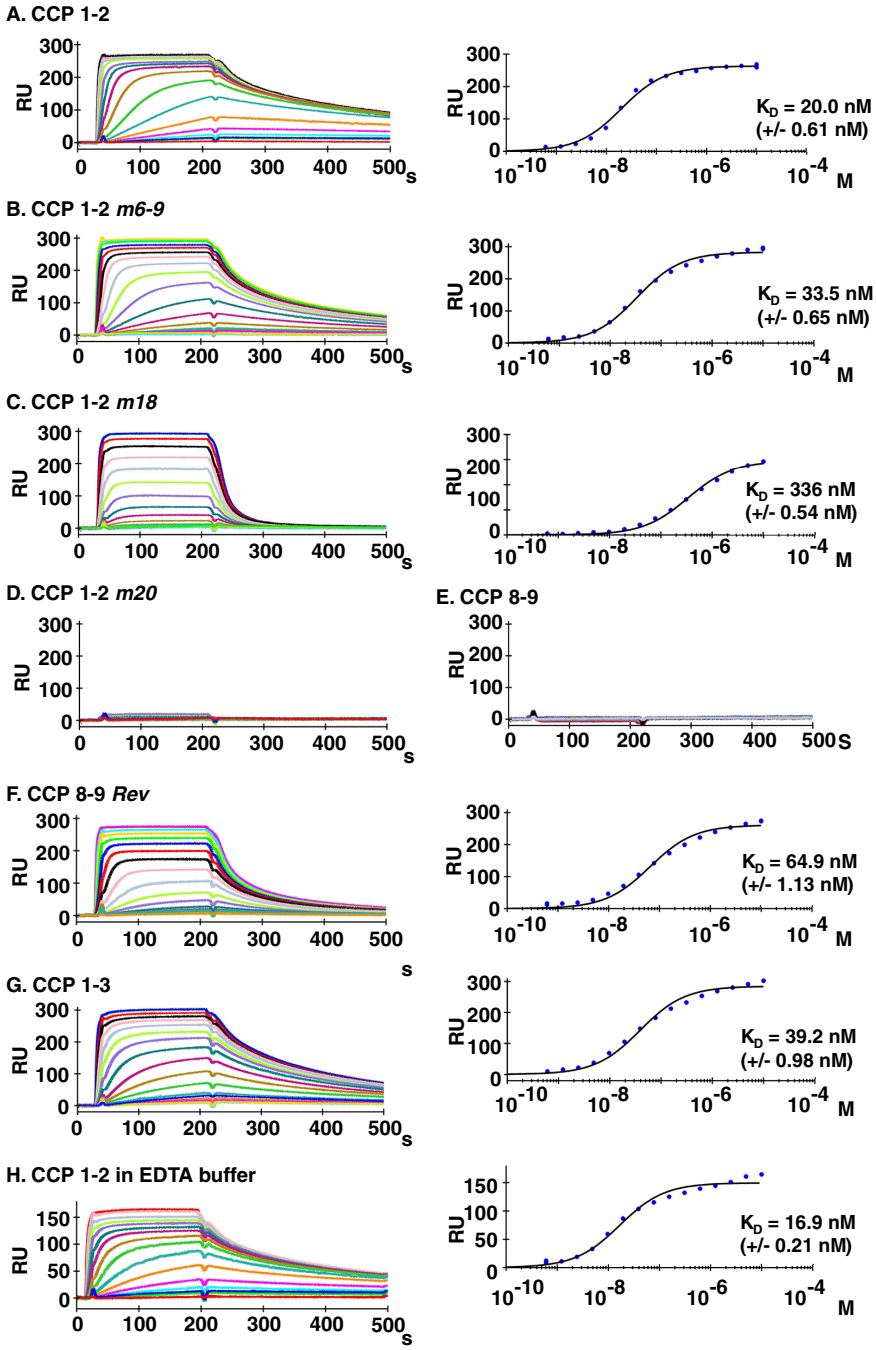


FIGURE 8. **Binding activity of CR1 constructs for rPfRh4 as measured by SPR.** *A*, 2900 resonance units (*RU*) rPfRh4 were coupled covalently to a carboxymethyldextran hydrogel biosensor chip. Reference subtracted sensorgrams of a 2-fold analyte dilution series for CCPs 1 and 2 binding to rPfRh4 are shown. The *right plot* is the respective response at steady state *versus* molar concentration fitted to a 1:1 steady state affinity model. CCPs 1 and 2 *m6–9* (*B*), CCPs 1 and 2 *m18* (*C*), CCPs 1 and 2 *m20* (*D*), CCPs 8 and 9 (*E*), and CCPs 8 and 9 *Rev* (*F*), are as described above. *G*, 1700 resonance units resonance units of rPfRh4 were coupled covalently to a carboxymethyldextran hydrogel biosensor chip, and a 2-fold concentration series of CCPs 1–3 in running buffer containing 1 mM $MgCl_2$ was probed for binding. Reference-subtracted sensorgrams are shown. The corresponding *right plots* are of responses achieved in *panel G* at steady state *versus* molar concentration fitted to a 1:1 steady state affinity model. *H*, binding was assayed for CCPs 1 and 2 in running buffer containing 3 mM EDTA $MgCl_2$ over a chip immobilized with 1950 resonance units of rPfRh4. A 2-fold concentration series was probed for binding, and reference subtracted sensorgrams are shown. The *right-hand graph* shows a plot of responses achieved in *panel H* at steady state *versus* molar concentration, which was fitted to a 1:1 steady state affinity model.

sible reason for this discrepancy may be that the novel purification method used for rPfRh4 produces a protein that more closely reflects the physiological conformation of the parasite adhesin on merozoites. Another reason may lie in a potential contribution of CCP domains 10–14 in binding to PfRh4. However, this is unlikely as the addition of 5H12 mAb is sufficient to prevent the formation of the complex, suggesting that CCPs 8 and 9 *Rev* and CCPs 1 and 2 essentially bind in the same mode to the exact same epitope on PfRh4. Furthermore, CCPs 8 and 9 *Rev* is able to block PfRh4 binding to red blood cells and functions as an inhibitor of *P. falciparum* PfRh4-CR1 invasion pathway to a similar extent as CCPs 1 and 2, CCPs 1–3, and sCR1.

TABLE 2

Characterization of anti-PfRh4 monoclonal antibodies

+ and − refer to positive and negative reactivity to PfRh4, respectively, using different detection techniques. For the interaction and functional assays, N, E, and I refer to neutral, enhancing, and inhibitory effects, respectively. WB, Western blot; IP, immunoprecipitation; IFA, indirect immunofluorescence assay; GIA, growth inhibition assay.

	Reactivity to PfRh4				PfRh4-CR1 Interaction and Functional assays						
mAb	ELISA	WB	IP	IFA	FRET	ELISA	IP	RBC binding	GIA	Epitope mapping	Affinity for Rh4
1A6	+	+	-	+	N	N		N	N	238-257	
1D11	+	+	-	+	N	N		E	N	Not found	
1H3	+	+	-	+	N	N		N	N	Not found	
2A1	+	+	-	+	N	N		N	N	522-545	
3A6	+	+	-	+	N	N		N	N	238-257	
5D2	+	+	-	+	N	N		N	N	266-281	
5H12	+	-	+	+	I	I	I	I	I	Not found	1.34 nM
6A5	+	+	+	+	N	E		~E	N	666-681	
6F12	+	+	-	-	N	N		N	N	718-737	
10C9	+	+	+	+	N	E	N	E	~E	662-681	0.48 nM

Our results highlight distinct regions within PfRh4 and CR1 that are important for mediating entry of *P. falciparum* parasites. We have demonstrated that it is possible to generate anti-PfRh4 neutralizing antibodies and a repertoire of CR1-based inhibitors that both block PfRh4-CR1 interaction and show functional inhibition in *P. falciparum* growth. Furthermore, we identified distinct epitopes within PfRh4 that should not be included in future vaccinations because of their ability to generate antibodies that enhance the interaction with CR1. This collection of anti-PfRh4 mAbs and CR1 constructs will be invaluable tools in the efforts to obtain an atomic resolution structure of the binding interface between PfRh4 and CR1 that will further inform the rational design of potent inhibitors.

Author Contributions—N. T. Y. L. and W.-H. T. contributed to project concept, acquisition, analyses, and interpretation of the data for the whole manuscript. C. Q. S. and M. J. H. expressed FH fragments in *P. pastoris*. C. Q. S., M. J. H., and C. S. L. performed surface plasmon resonance experiments. M. J. C. assisted with the design and experiments with FRET assay. A. T. K. performed *P. falciparum* growth assays. C. W. assisted in protein purification of rPfRh4 for antibody immunizations. A. F. C. also assisted in the design of experiments and interpretation of results, and all other authors were involved in writing and reviewing the manuscript.

Acknowledgments—We acknowledge the Victorian State Government Operational Infrastructure Support and Australian Government, National Health and Medical Research Council (NHMRC) Independent Research Institute Infrastructure Support Scheme (IRIISS). We acknowledge the excellent service at the Core Unit Mass Spectrometry and Proteomics (CUMP) of the University Clinic Ulm and thank Remco Sprangers for kind help in acquiring the ^{1}H NMR spectra.

References

1. Tham, W.-H., Healer, J., and Cowman, A. F. (2012) Erythrocyte and reticulocyte binding-like proteins of *Plasmodium falciparum*. *Trends Parasitol.* **28,** 23–30
2. Miller, L. H., Baruch, D. I., Marsh, K., and Doumbo, O. K. (2002) The pathogenic basis of malaria. *Nature* **415,** 673–679
3. Dvorak, J. A., Miller, L. H., Whitehouse, W. C., and Shiroishi, T. (1975) Invasion of erythrocytes by malaria merozoites. *Science* **187,** 748–750
4. Gilson, P. R., and Crabb, B. S. (2009) Morphology and kinetics of the three distinct phases of red blood cell invasion by *Plasmodium falciparum* merozoites. *Int. J. Parasitol.* **39,** 91–96
5. Miller, L. H., Aikawa, M., Johnson, J. G., and Shiroishi, T. (1979) Interaction between cytochalasin B-treated malarial parasites and erythrocytes: attachment and junction formation. *J. Exp. Med.* **149,** 172–184
6. Riglar, D. T., Richard, D., Wilson, D. W., Boyle, M. J., Dekiwadia, C., Turnbull, L., Angrisano, F., Marapana, D. S., Rogers, K. L., Whitchurch, C. B., Beeson, J. G., Cowman, A. F., Ralph, S. A., and Baum, J. (2011) Super-resolution dissection of coordinated events during malaria parasite invasion of the human erythrocyte. *Cell Host Microbe* **9,** 9–20
7. Aikawa, M., Miller, L. H., Johnson, J., and Rabbege, J. (1978) Erythrocyte entry by malarial parasites: a moving junction between erythrocyte and parasite. *J. Cell Biol.* **77,** 72–82
8. Adams, J. H., Sim, B. K., Dolan, S. A., Fang, X., Kaslow, D. C., and Miller, L. H. (1992) A family of erythrocyte binding proteins of malaria parasites. *Proc. Natl. Acad. Sci. U.S.A.* **89,** 7085–7089
9. Haynes, J. D., Dalton, J. P., Klotz, F. W., McGinniss, M. H., Hadley, T. J., Hudson, D. E., and Miller, L. H. (1988) Receptor-like specificity of a *Plasmodium knowlesi* malarial protein that binds to Duffy antigen ligands on erythrocytes. *J. Exp. Med.* **167,** 1873–1881
10. Triglia, T., Thompson, J., Caruana, S. R., Delorenzi, M., Speed, T., and Cowman, A. F. (2001) Identification of proteins from *Plasmodium falciparum* that are homologous to reticulocyte binding proteins in *Plasmodium vivax*. *Infect. Immun.* **69,** 1084–1092
11. Rayner, J. C., Vargas-Serrato, E., Huber, C. S., Galinski, M. R., and Barnwell, J. W. (2001) A *Plasmodium falciparum* homologue of *Plasmodium vivax* reticulocyte binding protein (PvRBP1) defines a trypsin-resistant erythrocyte invasion pathway. *J. Exp. Med.* **194,** 1571–1581
12. Hayton, K., Gaur, D., Liu, A., Takahashi, J., Henschen, B., Singh, S., Lambert, L., Furuya, T., Bouttenot, R., Doll, M., Nawaz, F., Mu, J., Jiang, L., Miller, L. H., and Wellems, T. E. (2008) Erythrocyte binding protein PfRH5 polymorphisms determine species-specific pathways of *Plasmodium falciparum* invasion. *Cell Host Microbe* **4,** 40–51
13. Lobo, C.-A., Rodriguez, M., Reid, M., and Lustigman, S. (2003) Glycophorin C is the receptor for the *Plasmodium falciparum* erythrocyte binding ligand PfEBP-2 (baebl). *Blood* **101,** 4628–4631

14. Maier, A. G., Duraisingh, M. T., Reeder, J. C., Patel, S. S., Kazura, J. W., Zimmerman, P. A., and Cowman, A. F. (2003) *Plasmodium falciparum* erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. *Nat. Med.* **9,** 87–92
15. Mayer, D. C., Mu, J.-B., Feng, X., Su, X. Z., and Miller, L. H. (2002) Polymorphism in a *Plasmodium falciparum* erythrocyte-binding ligand changes its receptor specificity. *J. Exp. Med.* **196,** 1523–1528
16. Sim, B. K., Chitnis, C. E., Wasniowska, K., Hadley, T. J., and Miller, L. H. (1994) Receptor and ligand domains for invasion of erythrocytes by *Plasmodium falciparum*. *Science* **264,** 1941–1944
17. Tham, W.-H., Wilson, D. W., Lopaticki, S., Schmidt, C. Q., Tetteh-Quarcoo, P. B., Barlow, P. N., Richard, D., Corbin, J. E., Beeson, J. G., and Cowman, A. F. (2010) Complement receptor 1 is the host erythrocyte receptor for *Plasmodium falciparum* PfRh4 invasion ligand. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 17327–17332
18. Crosnier, C., Bustamante, L. Y., Bartholdson, S. J., Bei, A. K., Theron, M., Uchikawa, M., Mboup, S., Ndir, O., Kwiatkowski, D. P., Duraisingh, M. T., Rayner, J. C., and Wright, G. J. (2011) Basigin is a receptor essential for erythrocyte invasion by *Plasmodium falciparum*. *Nature* **480,** 534–537
19. Spadafora, C., Awandare, G. A., Kopydlowski, K. M., Czege, J., Moch, J. K., Finberg, R. W., Tsokos, G. C., and Stoute, J. A. (2010) Complement receptor 1 is a sialic acid-independent erythrocyte receptor of *Plasmodium falciparum*. *PLoS Pathog.* **6,** e1000968
20. Richards, J. S., Arumugam, T. U., Reiling, L., Healer, J., Hodder, A. N., Fowkes, F. J., Cross, N., Langer, C., Takeo, S., Uboldi, A. D., Thompson, J. K., Gilson, P. R., Coppel, R. L., Siba, P. M., King, C. L., Torii, M., Chitnis, C. E., Narum, D. L., Mueller, I., Crabb, B. S., Cowman, A. F., Tsuboi, T., and Beeson, J. G. (2013) Identification and prioritization of merozoite antigens as targets of protective human immunity to *Plasmodium falciparum* malaria for vaccine and biomarker development. *J. Immunol.* **191,** 795–809
21. Persson, K. E., Fowkes, F. J., McCallum, F. J., Gicheru, N., Reiling, L., Richards, J. S., Wilson, D. W., Lopaticki, S., Cowman, A. F., Marsh, K., and Beeson, J. G. (2013) Erythrocyte-binding antigens of *Plasmodium falciparum* are targets of human inhibitory antibodies and function to evade naturally acquired immunity. *J. Immunol.* **191,** 785–794
22. Lopaticki, S., Maier, A. G., Thompson, J., Wilson, D. W., Tham, W.-H., Triglia, T., Gout, A., Speed, T. P., Beeson, J. G., Healer, J., and Cowman, A. F. (2011) Reticulocyte and erythrocyte binding-like proteins function cooperatively in invasion of human erythrocytes by malaria parasites. *Infect. Immun.* **79,** 1107–1117
23. Douglas, A. D., Williams, A. R., Knuepfer, E., Illingworth, J. J., Furze, J. M., Crosnier, C., Choudhary, P., Bustamante, L. Y., Zakutansky, S. E., Awuah, D. K., Alanine, D. G., Theron, M., Worth, A., Shimkets, R., Rayner, J. C., Holder, A. A., Wright, G. J., and Draper, S. J. (2014) Neutralization of *Plasmodium falciparum* merozoites by antibodies against PfRH5. *J. Immunol.* **192,** 245–258
24. Reiling, L., Richards, J. S., Fowkes, F. J., Wilson, D. W., Chokejindachai, W., Barry, A. E., Tham, W.-H., Stubbs, J., Langer, C., Donelson, J., Michon, P., Tavul, L., Crabb, B. S., Siba, P. M., Cowman, A. F., Mueller, I., and Beeson, J. G. (2012) The *Plasmodium falciparum* erythrocyte invasion ligand Pfrh4 as a target of functional and protective human antibodies against malaria. *PLoS ONE* **7,** e45253
25. Reiling, L., Richards, J. S., Fowkes, F. J., Barry, A. E., Triglia, T., Chokejindachai, W., Michon, P., Tavul, L., Siba, P. M., Cowman, A. F., Mueller, I., and Beeson, J. G. (2010) Evidence that the erythrocyte invasion ligand PfRh2 is a target of protective immunity against *Plasmodium falciparum* malaria. *J. Immunol.* **185,** 6157–6167
26. Gao, X., Gunalan, K., Yap, S. S., and Preiser, P. R. (2013) Triggers of key calcium signals during erythrocyte invasion by *Plasmodium falciparum*. *Nat. Commun.* **4,** 2862
27. Douglas, A. D., Baldeviano, G. C., Lucas, C. M., Lugo-Roman, L. A., Crosnier, C., Bartholdson, S. J., Diouf, A., Miura, K., Lambert, L. E., Ventocilla, J. A., Leiva, K. P., Milne, K. H., Illingworth, J. J., Spencer, A. J., Hjerrild, K. A., Alanine, D. G., Turner, A. V., Moorhead, J. T., Edgel, K. A., Wu, Y., Long, C. A., Wright, G. J., Lescano, A. G., and Draper, S. J. (2015) A PfRH5-based vaccine is efficacious against heterologous strain blood-stage *Plasmodium falciparum* infection in *Aotus* monkeys. *Cell Host Microbe.* **17,** 130–139
28. Chen, L., Xu, Y., Healer, J., Thompson, J. K., Smith, B. J., Lawrence, M. C., and Cowman, A. F. (2014) Crystal structure of PfRh5, an essential *P. falciparum* ligand for invasion of human erythrocytes. *Elife* 10.7554/eLife.04187
29. Wright, K. E., Hjerrild, K. A., Bartlett, J., Douglas, A. D., Jin, J., Brown, R. E., Illingworth, J. J., Ashfield, R., Clemmensen, S. B., de Jongh, W. A., Draper, S. J., and Higgins, M. K. (2014) Structure of malaria invasion protein RH5 with erythrocyte basigin and blocking antibodies. *Nature* **515,** 427–430
30. Schmidt, C. Q., Kennedy, A. T., and Tham, W.-H. (2015) More than just immune evasion: hijacking complement by *Plasmodium falciparum*. *Mol. Immunol.* **67,** 71–84
31. Tham, W.-H., Wilson, D. W., Reiling, L., Chen, L., Beeson, J. G., and Cowman, A. F. (2009) Antibodies to reticulocyte binding protein-like homologue 4 inhibit invasion of *Plasmodium falciparum* into human erythrocytes. *Infect. Immun.* **77,** 2427–2435
32. Park, H. J., Guariento, M., Maciejewski, M., Hauhart, R., Tham, W.-H., Cowman, A. F., Schmidt, C. Q., Mertens, H. D., Liszewski, M. K., Hourcade, D. E., Barlow, P. N., and Atkinson, J. P. (2014) Using mutagenesis and structural biology to map the binding site for the *Plasmodium falciparum* merozoite protein PfRh4 on the human immune adherence receptor. *J. Biol. Chem.* **289,** 450–463
33. Tham, W.-H., Schmidt, C. Q., Hauhart, R. E., Guariento, M., Tetteh-Quarcoo, P. B., Lopaticki, S., Atkinson, J. P., Barlow, P. N., and Cowman, A. F. (2011) *Plasmodium falciparum* uses a key functional site in complement receptor type-1 for invasion of human erythrocytes. *Blood* **118,** 1923–1933
34. Awandare, G. A., Spadafora, C., Moch, J. K., Dutta, S., Haynes, J. D., and Stoute, J. A. (2011) *Plasmodium falciparum* field isolates use complement receptor 1 (CR1) as a receptor for invasion of erythrocytes. *Mol. Biochem. Parasitol.* **177,** 57–60
35. Weiss, G. E., Gilson, P. R., Taechalertpaisarn, T., Tham, W.-H., de Jong, N. W., Harvey, K. L., Fowkes, F. J., Barlow, P. N., Rayner, J. C., Wright, G. J., Cowman, A. F., and Crabb, B. S. (2015) Revealing the sequence and resulting cellular morphology of receptor-ligand interactions during *Plasmodium falciparum* invasion of erythrocytes. *PLoS Pathog.* **11,** e1004670
36. Boyle, M. J., Wilson, D. W., Richards, J. S., Riglar, D. T., Tetteh, K. K., Conway, D. J., Ralph, S. A., Baum, J., and Beeson, J. G. (2010) Isolation of viable *Plasmodium falciparum* merozoites to define erythrocyte invasion events and advance vaccine and drug development. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 14378–14383
37. Wilson, D. W., Crabb, B. S., and Beeson, J. G. (2010) Development of fluorescent *Plasmodium falciparum* for *in vitro* growth inhibition assays. *Malar. J.* **9,** 152
38. Schmidt, C. Q., Slingsby, F. C., Richards, A., and Barlow, P. N. (2011) Production of biologically active complement factor H in therapeutically useful quantities. *Protein Expr. Purif.* **76,** 254–263
39. Triglia, T., Tham, W.-H., Hodder, A., and Cowman, A. F. (2009) Reticulocyte binding protein homologues are key adhesins during erythrocyte invasion by *Plasmodium falciparum*. *Cell. Microbiol.* **11,** 1671–1687

Characterization of Inhibitors and Monoclonal Antibodies That Modulate the Interaction between *Plasmodium falciparum* Adhesin PfRh4 with Its Erythrocyte Receptor Complement Receptor 1

Nicholas T. Y. Lim, Markus J. Harder, Alexander T. Kennedy, Clara S. Lin, Christopher Weir, Alan F. Cowman, Melissa J. Call, Christoph Q. Schmidt and Wai-Hong Tham

J. Biol. Chem. 2015, 290:25307-25321.
doi: 10.1074/jbc.M115.657171 originally published online August 31, 2015